

A HUMAN PROTEIN KINASE DOMAIN-CONTAINING PROTEIN

CROSS REFERENCE TO RELATED APPLICATIONS

5 This application claims priority under 35 U.S.C. § 365(c) to international patent application no. PCT/US01/00665, PCT/US01/00667, and PCT/US01/00668, all filed January 30, 2001; claims priority under 35 U.S.C. § 120 to commonly owned and copending U.S. application serial no. 09/864,761, filed May 23,
10 2001; claims priority to United States provisional application serial no. 60/335,941, filed October 24, 2001; the disclosures of which are incorporated herein by reference in their entireties.

15 REFERENCE TO SEQUENCE LISTING SUBMITTED ON COMPACT DISC

 The present application includes a Sequence Listing filed on a single CD-R disc, provided in duplicate, containing a single file named pto_PB0179.txt, having 33 kilobytes, last
20 modified on January 17, 2002 and recorded January 24, 2002. The Sequence Listing contained in said file on said disc is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

25 The present invention relates to a novel human protein kinase. More specifically, the invention provides isolated nucleic acid molecules encoding a human serine/threonine/tyrosine kinase domain-containing protein
30 (STTK), fragments thereof, vectors and host cells comprising isolated nucleic acid molecules encoding STTK, STTK polypeptides, antibodies, transgenic cells and non-human organisms, and diagnostic, therapeutic, and investigational methods of using the same.

BACKGROUND OF THE INVENTION

In all eukaryotes, reversible protein phosphorylation
5 plays a major role in regulating basic cellular functions such
as DNA replication, gene transcription, protein translation,
cell cycle control, and metabolism. Protein phosphorylation is
also important for more advanced functions in higher eukaryotes
such as cell and organ differentiation, cell-cell
10 communication, cell survival, and synaptic transmission.

Protein phosphorylation is mediated by protein
kinases, enzymes that catalyze the transfer of the γ -phosphate
of ATP to protein substrates. Protein kinases are critical for
signal amplification and distribution because a single protein
15 kinase can phosphorylate many different target proteins and
therefore modify their activities. Furthermore, some of those
targets can themselves be protein kinases and leads to further
rounds of signal amplification and distribution. Due to the key
roles protein kinases play in cellular signaling processes,
20 their catalytic activity is tightly controlled in normal cells
by protein phosphatases, by other protein kinases, and by
autoregulatory mechanisms. Hubbard et al., *J. Biol. Chem.*
273:11987-11990 (1998).

There are two major groups of protein kinases, the
25 protein tyrosine kinases (PTKs) that catalyze the transfer of
the γ -phosphate of ATP to tyrosine residues of protein
substrates, and the protein serine/threonine kinases that
catalyze the transfer of the γ -phosphate of ATP to serine or
threonine residues of protein substrates.

30 The PTKs can be further divided into two subfamilies,
receptor tyrosine kinases (RTKs) and non-receptor tyrosine
kinases (NRTKs). RTKs contain a transmembrane portion, which
spans the plasma membrane, an extracellular portion, which

binds ligand, and an intracellular portion, which possesses catalytic and regulatory domains. The RTK family includes the insulin receptor and the receptors for many growth factors such as epidermal (EGF), platelet-derived (PDGF), fibroblast (FGF), and nerve growth factors. NRTKs contain no extracellular or transmembrane portion but possess modular domains that are responsible for subcellular targeting and regulation of catalytic activity. The NRTK family includes Src, Abl, FAK, and the JAKs among many others. The serine/threonine kinases also are consisted of membrane bound kinases such as activin receptor and non-membrane bound kinases such as cAMP activated kinase. Ullrich et al., *Cell* 61:203-212 (1990); Schlessinger et al., *Cell* 103:211-225 (2000).

The activities of protein kinases are under tight regulation. The phosphorylation of a protein kinase itself often serves to regulate its catalytic activity. An exposed loop near the active site of the kinase catalytic domain, termed the activation loop, is often a major site on kinases for regulatory phosphorylation. Many kinases require one or two amino acids in this loop to be phosphorylated for efficient catalytic activity. This mechanism allows inter-kinase regulation and forms the basis for the kinase cascades that are common in eukaryotic cells. The kinase activity is also regulated by the use of an auto-inhibitory segment attached in cis to the protein kinase catalytic domain. For the calmodulin-dependent kinases, a conserved region C-terminal to the canonical kinase domain folds back onto the catalytic domain perturbing the ATP-binding site. Ca²⁺-bound calmodulin activates the kinase by binding the regulatory region and relieving the kinase domain of this inhibitory interaction. Yang et al., *J. Biol. Chem.* 274:26199-26208 (1999); Kolodziej et al., *J. Biol. Chem.* 275(19):14354-14359 (2000). The use of an inhibitory domain provides a mechanism for a unique apical

signal to trigger a kinase cascade. This signal can then be propagated and amplified by a kinase phosphorylation cascade, resulting in the appropriate physiological response.

Aberrant protein phosphorylation is a common cause of cancer and other human diseases. Therefore, protein kinases are considered excellent drug target molecules for drug discovery. For example, chronic myeloid leukemia (CML) is a stem cell disorder resulting from a 'reciprocal translocation' between chromosomes 9 and 22, resulting in the expression of Bcr-Abl, an abnormal protein tyrosine kinase that causes the uncontrolled proliferation of white blood cells. GleevecTM (Imatinib mesylate) is a chemical compound designed to inhibit the functions of Bcr-Abl. It was shown to inhibit proliferation and induce apoptosis in Bcr-Abl positive cells. O'Dwyer and Druker, *J. Intern. Med.* 250:3-9 (2001). In May 2001, the Food and Drug Administration approved the use of GleevecTM for treatment of CML.

Due to the important roles protein kinases play in mediating the signaling pathways of cellular activities, and in the therapy as well as diagnosis of cancer, it is desirable to identify and to characterize additional protein kinases.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing isolated nucleic acids that encode a human serine/threonine/tyrosine protein kinase domain-containing protein (STTK), and fragments thereof.

In other aspects, the invention provides vectors for propagating and expressing the nucleic acids of the present invention, host cells comprising the nucleic acids and vectors of the present invention, proteins, protein fragments, and protein fusions of the human STTK, and antibodies thereto.

The invention further provides pharmaceutical formulations of the nucleic acids, proteins, and antibodies of the present invention.

5 In other aspects, the invention provides transgenic cells and non-human organisms comprising STTK nucleic acids, and transgenic cells and non-human organisms with targeted disruption of the endogenous orthologue of the human STTK.

The invention additionally provides diagnostic, investigational, and therapeutic methods based on the human
10 STTK nucleic acids, proteins, and antibodies of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

15 The above and other objects and advantages of the present invention will be apparent upon consideration of the following detailed description taken in conjunction with the accompanying drawings, in which like characters refer to like parts throughout, and in which:

20 FIG. 1 (A) schematizes the protein domain structure of human STTK, FIG. 1 (B) shows the alignment of the protein kinase domain of STTK with that of other protein kinases;

FIG. 2 is a map showing the genomic structure of human STTK encoded at chromosome 1q32.1;

25 FIG. 3 presents the nucleotide and predicted amino acid sequences of human STTK; and

FIG. 4 presents the expression profile of STTK by RT-PCR analysis.

30 DETAILED DESCRIPTION OF THE INVENTION

Mining the sequence of the human genome for novel human genes, the present inventors have identified human STTK, a protein kinase domain-containing protein.

As schematized in FIG. 1, the newly isolated gene product shares a key protein domain, the kinase domain, with other protein kinases. The shared structural feature strongly implies that human STTK plays a role similar to that of other protein kinases in signal transduction in the cell.

Like other protein kinases, STTK has a protein kinase domain. In STTK, the kinase domain occurs at residues 653 - 897 (<http://pfam.wustl.edu/>). Protein kinase domain is the catalytic domain of a protein kinase that catalyzes the transfer of the γ -phosphate of ATP to protein substrates. Protein kinases can be divided into two major groups, the serine/threonine kinases and the tyrosine kinases, depending on their substrate specificities. Since the kinase domain of STTK is similarly homologous to serine/threonine kinase domain and tyrosine kinase domain, STTK is predicted to be able to phosphorylate serine/threonine as well as tyrosine residues in its substrates.

Other signatures of the STTK protein were identified by searching the PROSITE database (<http://www.expasy.ch/tools/scnpsit1.html>). These include two N-glycosylation sites (60-63, 594-597), fifteen Protein kinase C phosphorylation sites, nine Casein kinase II phosphorylation sites, two Tyrosine kinase phosphorylation site (305-312 and 395-402), and fifteen N-myristoylation sites.

FIG. 2 shows the genomic organization of human STTK.

At the top is shown the one bacterial artificial chromosome (BAC), with GenBank accession number (AC018711.4), that spans the STTK locus. The genome-derived single-exon probes first used to demonstrate expression from this locus are shown below the BAC and labeled "516", "578", and "550",

respectively. The 516 bp probe includes sequence drawn from exon 3, with additional sequence from intron 3. The 578 bp probe includes sequence drawn from exon 8, with additional sequence from intron 7 and intron 8. The 550 bp probe includes
5 sequence drawn from exon 13, with additional sequence from intron 12 as well as sequence 3' of exon 13.

As shown in FIG. 2, STTK encodes a protein of 929 amino acids and is comprised of exons 1 - 13. STTK has a predicted molecular weights, prior to any post-translational
10 modification of 105.3 kD.

As further discussed in the examples herein, expression of STTK was assessed using hybridization to genome-derived single exon microarrays and RT-PCR. Microarray analysis of the exons three, eight, and thirteen showed
15 expression in adult liver, bone marrow, brain, fetal liver, kidney, lung, and placenta. RT-PCR confirmed microarray data, and further provided expression data for colon, heart, skeletal muscle as well as a cell line, HeLa.

As more fully described below, the present invention provides isolated nucleic acids that encode human STTK and fragments thereof. The invention further provides vectors for propagation and expression of the nucleic acids of the present invention, host cells comprising the nucleic acids and vectors of the present invention, proteins, protein fragments, and
25 protein fusions of the present invention, and antibodies specific for all or any one of the isoforms. The invention provides pharmaceutical formulations of the nucleic acids, proteins, and antibodies of the present invention. The invention further provides transgenic cells and non-human
30 organisms comprising STTK nucleic acids, and transgenic cells and non-human organisms with targeted disruption of the endogenous orthologue of the STTK. The invention additionally provides diagnostic, investigational, and therapeutic methods

based on the STTK nucleic acids, proteins, and antibodies of the present invention.

DEFINITIONS

5

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by one of ordinary skill in the art to which this invention belongs.

10

As used herein, **"nucleic acid"** (synonymously, **"polynucleotide"**) includes polynucleotides having natural nucleotides in native 5'-3' phosphodiester linkage - e.g., DNA or RNA - as well as polynucleotides that have nonnatural nucleotide analogues, nonnative internucleoside bonds, or both, so long as the nonnatural polynucleotide is capable of sequence-discriminating basepairing under experimentally desired conditions. Unless otherwise specified, the term **"nucleic acid"** includes any topological conformation; the term thus explicitly comprehends single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular, and padlocked conformations.

15

20

As used herein, an **"isolated nucleic acid"** is a nucleic acid molecule that exists in a physical form that is nonidentical to any nucleic acid molecule of identical sequence as found in nature; **"isolated"** does not require, although it does not prohibit, that the nucleic acid so described has itself been physically removed from its native environment.

25

30

For example, a nucleic acid can be said to be **"isolated"** when it includes nucleotides and/or internucleoside bonds not found in nature. When instead composed of natural nucleosides in phosphodiester linkage, a nucleic acid can be said to be **"isolated"** when it exists at a purity not found in nature, where purity can be adjudged with respect to the

presence of nucleic acids of other sequence, with respect to the presence of proteins, with respect to the presence of lipids, or with respect the presence of any other component of a biological cell, or when the nucleic acid lacks sequence that
5 flanks an otherwise identical sequence in an organism's genome, or when the nucleic acid possesses sequence not identically present in nature.

As so defined, "isolated nucleic acid" includes nucleic acids integrated into a host cell chromosome at a
10 heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

As used herein, an isolated nucleic acid "encodes" a reference polypeptide when at least a portion of the nucleic
15 acid, or its complement, can be directly translated to provide the amino acid sequence of the reference polypeptide, or when the isolated nucleic acid can be used, alone or as part of an expression vector, to express the reference polypeptide *in vitro*, in a prokaryotic host cell, or in a eukaryotic host
20 cell.

As used herein, the term "**exon**" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute contiguous sequence to a mature mRNA transcript.

As used herein, the phrase "**open reading frame**" and the equivalent acronym "**ORF**" refer to that portion of a
25 transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly
30 divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "**ORF-encoded peptide**" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "**degenerate variant**" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to
5 that translated from the reference nucleic acid sequence.

As used herein, the term "**microarray**" and the equivalent phrase "**nucleic acid microarray**" refer to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids
10 being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed.

As so defined, the term "microarray" and phrase "nucleic acid microarray" include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical
15 Approach Series), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1 - 60 (1999); and Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376), the disclosures of which are incorporated herein
20 by reference in their entireties.

As so defined, the term "microarray" and phrase "nucleic acid microarray" also include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are distributably disposed on a plurality of
25 beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner et al., *Proc. Natl. Acad. Sci. USA* 97(4):166501670 (2000), the disclosure of which is incorporated herein by reference in its entirety; in such case, the term "microarray" and phrase "nucleic acid microarray"
30 refer to the plurality of beads in aggregate.

As used herein with respect to solution phase hybridization, the term "**probe**", or equivalently, "**nucleic acid probe**" or "**hybridization probe**", refers to an isolated nucleic

acid of known sequence that is, or is intended to be,
detectably labeled. As used herein with respect to a nucleic
acid microarray, the term "probe" (or equivalently "nucleic
acid probe" or "hybridization probe") refers to the isolated
5 nucleic acid that is, or is intended to be, bound to the
substrate. In either such context, the term "target" refers to
nucleic acid intended to be bound to probe by sequence
complementarity.

As used herein, the expression **"probe comprising SEQ
10 ID NO:X"**, and variants thereof, intends a nucleic acid probe,
at least a portion of which probe has either (i) the sequence
directly as given in the referenced SEQ ID NO:X, or (ii) a
sequence complementary to the sequence as given in the
referenced SEQ ID NO:X, the choice as between sequence directly
15 as given and complement thereof dictated by the requirement
that the probe be complementary to the desired target.

As used herein, the phrases **"expression of a probe"**
and **"expression of an isolated nucleic acid"** and their
linguistic equivalents intend that the probe or, (respectively,
20 the isolated nucleic acid), or a probe (or, respectively,
isolated nucleic acid) complementary in sequence thereto, can
hybridize detectably under high stringency conditions to a
sample of nucleic acids that derive from mRNA transcripts from
a given source. For example, and by way of illustration only,
25 expression of a probe in "liver" means that the probe can
hybridize detectably under high stringency conditions to a
sample of nucleic acids that derive from mRNA obtained from
liver.

As used herein, **"a single exon probe"** comprises at
30 least part of an exon ("reference exon") and can hybridize
detectably under high stringency conditions to
transcript-derived nucleic acids that include the reference
exon. The single exon probe will not, however, hybridize

detectably under high stringency conditions to nucleic acids that lack the reference exon and that consist of one or more exons that are found adjacent to the reference exon in the genome.

5 For purposes herein, **"high stringency conditions"** are defined for solution phase hybridization as aqueous hybridization (*i.e.*, free of formamide) in 6X SSC (where 20X SSC contains 3.0 M NaCl and 0.3 M sodium citrate), 1% SDS at 65°C for at least 8 hours, followed by one or more washes in
10 0.2X SSC, 0.1% SDS at 65°C. **"Moderate stringency conditions"** are defined for solution phase hybridization as aqueous hybridization (*i.e.*, free of formamide) in 6X SSC, 1% SDS at 65°C for at least 8 hours, followed by one or more washes in 2x SSC, 0.1% SDS at room temperature.

15 For microarray-based hybridization, standard "high stringency conditions" are defined as hybridization in 50% formamide, 5X SSC, 0.2 µg/µl poly(dA), 0.2 µg/µl human cot1 DNA, and 0.5% SDS, in a humid oven at 42°C overnight, followed by successive washes of the microarray in 1X SSC, 0.2% SDS at
20 55°C for 5 minutes, and then 0.1X SSC, 0.2% SDS, at 55°C for 20 minutes. For microarray-based hybridization, "moderate stringency conditions", suitable for cross-hybridization to mRNA encoding structurally- and functionally-related proteins, are defined to be the same as those for high stringency
25 conditions but with reduction in temperature for hybridization and washing to room temperature (approximately 25°C).

 As used herein, the terms **"protein"**, **"polypeptide"**, and **"peptide"** are used interchangeably to refer to a naturally-occurring or synthetic polymer of amino acid monomers
30 (residues), irrespective of length, where amino acid monomer here includes naturally-occurring amino acids, naturally-occurring amino acid structural variants, and synthetic non-

naturally occurring analogs that are capable of participating in peptide bonds. The terms "protein", "polypeptide", and "peptide" explicitly permits of post-translational and post-synthetic modifications, such as glycosylation.

5 The term **"oligopeptide"** herein denotes a protein, polypeptide, or peptide having 25 or fewer monomeric subunits.

 The phrases **"isolated protein"**, **"isolated polypeptide"**, **"isolated peptide"** and **"isolated oligopeptide"** refer to a protein (or respectively to a polypeptide, peptide, 10 or oligopeptide) that is nonidentical to any protein molecule of identical amino acid sequence as found in nature; "isolated" does not require, although it does not prohibit, that the protein so described has itself been physically removed from its native environment.

15 For example, a protein can be said to be "isolated" when it includes amino acid analogues or derivatives not found in nature, or includes linkages other than standard peptide bonds.

 When instead composed entirely of natural amino acids 20 linked by peptide bonds, a protein can be said to be "isolated" when it exists at a purity not found in nature — where purity can be adjudged with respect to the presence of proteins of other sequence, with respect to the presence of non-protein compounds, such as nucleic acids, lipids, or other components 25 of a biological cell, or when it exists in a composition not found in nature, such as in a host cell that does not naturally express that protein.

 A **"purified protein"** (equally, a purified polypeptide, peptide, or oligopeptide) is an isolated protein, 30 as above described, present at a concentration of at least 95%, as measured on a weight basis with respect to total protein in a composition. A **"substantially purified protein"** (equally, a substantially purified polypeptide, peptide, or oligopeptide)

is an isolated protein, as above described, present at a concentration of at least 70%, as measured on a weight basis with respect to total protein in a composition.

As used herein, the phrase "**protein isoforms**" refers to a plurality of proteins having nonidentical primary amino acid sequence but that share amino acid sequence encoded by at least one common exon.

As used herein, the phrase "**alternative splicing**" and its linguistic equivalents includes all types of RNA processing that lead to expression of plural protein isoforms from a single gene; accordingly, the phrase "**splice variant(s)**" and its linguistic equivalents embraces mRNAs transcribed from a given gene that, however processed, collectively encode plural protein isoforms. For example, and by way of illustration only, splice variants can include exon insertions, exon extensions, exon truncations, exon deletions, alternatives in the 5' untranslated region ("5' UT") and alternatives in the 3' untranslated region ("3' UT"). Such 3' alternatives include, for example, differences in the site of RNA transcript cleavage and site of poly(A) addition. See, e.g., Gautheret et al., *Genome Res.* 8:524-530 (1998).

As used herein, "**orthologues**" are separate occurrences of the same gene in multiple species. The separate occurrences have similar, albeit nonidentical, amino acid sequences, the degree of sequence similarity depending, in part, upon the evolutionary distance of the species from a common ancestor having the same gene.

As used herein, the term "**paralogues**" indicates separate occurrences of a gene in one species. The separate occurrences have similar, albeit nonidentical, amino acid sequences, the degree of sequence similarity depending, in part, upon the evolutionary distance from the gene duplication event giving rise to the separate occurrences.

As used herein, the term "**homologues**" is generic to "orthologues" and "paralogues".

As used herein, the term "**antibody**" refers to a polypeptide, at least a portion of which is encoded by at least one immunoglobulin gene, or fragment thereof, and that can bind specifically to a desired target molecule. The term includes naturally-occurring forms, as well as fragments and derivatives.

Fragments within the scope of the term "antibody" include those produced by digestion with various proteases, those produced by chemical cleavage and/or chemical dissociation, and those produced recombinantly, so long as the fragment remains capable of specific binding to a target molecule. Among such fragments are Fab, Fab', Fv, F(ab)'₂, and single chain Fv (scFv) fragments.

Derivatives within the scope of the term include antibodies (or fragments thereof) that have been modified in sequence, but remain capable of specific binding to a target molecule, including: interspecies chimeric and humanized antibodies; antibody fusions; heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies (see, e.g., Marasco (ed.), Intracellular Antibodies: Research and Disease Applications, Springer-Verlag New York, Inc. (1998) (ISBN: 3540641513), the disclosure of which is incorporated herein by reference in its entirety).

As used herein, antibodies can be produced by any known technique, including harvest from cell culture of native B lymphocytes, harvest from culture of hybridomas, recombinant expression systems, and phage display.

As used herein, "**antigen**" refers to a ligand that can be bound by an antibody; an antigen need not itself be

immunogenic. The portions of the antigen that make contact with the antibody are denominated "**epitopes**".

"**Specific binding**" refers to the ability of two molecular species concurrently present in a heterogeneous (inhomogeneous) sample to bind to one another in preference to binding to other molecular species in the sample. Typically, a specific binding interaction will discriminate over adventitious binding interactions in the reaction by at least two-fold, more typically by at least 10-fold, often at least 100-fold; when used to detect analyte, specific binding is sufficiently discriminatory when determinative of the presence of the analyte in a heterogeneous (inhomogeneous) sample. Typically, the affinity or avidity of a specific binding reaction is least about 10^{-7} M, with specific binding reactions of greater specificity typically having affinity or avidity of at least 10^{-8} M to at least about 10^{-9} M.

As used herein, "**molecular binding partners**" – and equivalently, "**specific binding partners**" – refer to pairs of molecules, typically pairs of biomolecules, that exhibit specific binding. Nonlimiting examples are receptor and ligand, antibody and antigen, and biotin to any of avidin, streptavidin, neutrAvidin and captAvidin.

The term "**antisense**", as used herein, refers to a nucleic acid molecule sufficiently complementary in sequence, and sufficiently long in that complementary sequence, as to hybridize under intracellular conditions to (i) a target mRNA transcript or (ii) the genomic DNA strand complementary to that transcribed to produce the target mRNA transcript.

The term "**portion**", as used with respect to nucleic acids, proteins, and antibodies, is synonymous with "fragment".

NUCLEIC ACID MOLECULES

In a first aspect, the invention provides isolated nucleic acids that encode human STTK, variants having at least
5 65% sequence identity thereto, degenerate variants thereof, variants that encode STTK proteins having conservative or moderately conservative substitutions, cross-hybridizing nucleic acids, and fragments thereof.

FIG. 3 presents the nucleotide sequence of the human
10 STTK cDNA clone, with predicted amino acid translation; the sequences are further presented in the Sequence Listing, incorporated herein by reference in its entirety, in SEQ ID NOs: 1 (full length nucleotide sequence of human STTK cDNA) and 3 (full length amino acid coding sequence of human STTK).

15 Unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the
20 given sequence intends ribonucleotides, with uridine substituted for thymidine.

Unless otherwise indicated, nucleotide sequences of the isolated nucleic acids of the present invention were determined by sequencing a DNA molecule that had resulted,
25 directly or indirectly, from at least one enzymatic polymerization reaction (e.g., reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Amersham Biosciences, Sunnyvale, CA, USA), or by reliance upon such sequence or upon genomic
30 sequence prior-accessioned into a public database. Unless otherwise indicated, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined.

As a consequence, any nucleic acid sequence presented herein may contain errors introduced by erroneous incorporation of nucleotides during polymerization, by erroneous base calling by the automated sequencer (although such sequencing errors have been minimized for the nucleic acids directly determined herein, unless otherwise indicated, by the sequencing of each of the complementary strands of a duplex DNA), or by similar errors accessioned into the public database. Such errors can readily be identified and corrected by resequencing of the genomic locus using standard techniques.

Single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes - more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409:860 - 921 (2001) - and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein.

Accordingly, it is an aspect of the present invention to provide nucleic acids not only identical in sequence to those described with particularity herein, but also to provide isolated nucleic acids at least about 65% identical in sequence to those described with particularity herein, typically at least about 70%, 75%, 80%, 85%, or 90% identical in sequence to those described with particularity herein, usefully at least about 91%, 92%, 93%, 94%, or 95% identical in sequence to those described with particularity herein, usefully at least about 96%, 97%, 98%, or 99% identical in sequence to those described with particularity herein, and, most conservatively, at least about 99.5%, 99.6%, 99.7%, 99.8% and 99.9% identical in sequence to those described with particularity herein. These

sequence variants can be naturally occurring or can result from human intervention, as by random or directed mutagenesis.

For purposes herein, percent identity of two nucleic acid sequences is determined using the procedure of Tatiana et al., "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", *FEMS Microbiol Lett.* 174:247-250 (1999), which procedure is effectuated by the computer program BLAST 2 SEQUENCES, available online at

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>.

To assess percent identity of nucleic acids, the BLASTN module of BLAST 2 SEQUENCES is used with default values of (i) reward for a match: 1; (ii) penalty for a mismatch: -2; (iii) open gap 5 and extension gap 2 penalties; (iv) gap X_dropoff 50 expect 10 word size 11 filter, and both sequences are entered in their entireties.

As is well known, the genetic code is degenerate, with each amino acid except methionine translated from a plurality of codons, thus permitting a plurality of nucleic acids of disparate sequence to encode the identical protein.

As is also well known, codon choice for optimal expression varies from species to species. The isolated nucleic acids of the present invention being useful for expression of STTK proteins and protein fragments, it is, therefore, another aspect of the present invention to provide isolated nucleic acids that encode STTK proteins and portions thereof not only identical in sequence to those described with particularity herein, but degenerate variants thereof as well.

As is also well known, amino acid substitutions occur frequently among natural allelic variants, with conservative substitutions often occasioning only *de minimis* change in protein function.

Accordingly, it is an aspect of the present invention to provide nucleic acids not only identical in sequence to

those described with particularity herein, but also to provide isolated nucleic acids that encode STTK, and portions thereof, having conservative amino acid substitutions, and also to provide isolated nucleic acids that encode STTK, and portions thereof, having moderately conservative amino acid substitutions.

Although there are a variety of metrics for calling conservative amino acid substitutions, based primarily on either observed changes among evolutionarily related proteins or on predicted chemical similarity, for purposes herein a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix reproduced herein below (see Gonnet et al., *Science* 256(5062):1443-5 (1992)):

15		A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
	A	2	-1	0	0	0	0	0	0	-1	-1	-1	0	-1	-2	0	1	1	-4	-2	0
	R	-1	5	0	0	-2	2	0	-1	1	-2	-2	3	-2	-3	-1	0	0	-2	-2	-2
	N	0	0	4	2	-2	1	1	0	1	-3	-3	1	-2	-3	-1	1	0	-4	-1	-2
	D	0	0	2	5	-3	1	3	0	0	-4	-4	0	-3	-4	-1	0	0	-5	-3	-3
20	C	0	-2	-2	-3	12	-2	-3	-2	-1	-1	-2	-3	-1	-1	-3	0	0	-1	0	0
	Q	0	2	1	1	-2	3	2	-1	1	-2	-2	2	-1	-3	0	0	0	-3	-2	-2
	E	0	0	1	3	-3	2	4	-1	0	-3	-3	1	-2	-4	0	0	0	-4	-3	-2
	G	0	-1	0	0	-2	-1	-1	7	-1	-4	-4	-1	-4	-5	-2	0	-1	-4	-4	-3
	H	-1	1	1	0	-1	1	0	-1	6	-2	-2	1	-1	0	-1	0	0	-1	2	-2
25	I	-1	-2	-3	-4	-1	-2	-3	-4	-2	4	3	-2	2	1	-3	-2	-1	-2	-1	3
	L	-1	-2	-3	-4	-2	-2	-3	-4	-2	3	4	-2	3	2	-2	-2	-1	-1	0	2
	K	0	3	1	0	-3	2	1	-1	1	-2	-2	3	-1	-3	-1	0	0	-4	-2	-2
	M	-1	-2	-2	-3	-1	-1	-2	-4	-1	2	3	-1	4	2	-2	-1	-1	-1	0	2
	F	-2	-3	-3	-4	-1	-3	-4	-5	0	1	2	-3	2	7	-4	-3	-2	4	5	0
30	P	0	-1	-1	-1	-3	0	0	-2	-1	-3	-2	-1	-2	-4	8	0	0	-5	-3	-2
	S	1	0	1	0	0	0	0	0	0	-2	-2	0	-1	-3	0	2	2	-3	-2	-1
	T	1	0	0	0	0	0	0	-1	0	-1	-1	0	-1	-2	0	2	2	-4	-2	0
	W	-4	-2	-4	-5	-1	-3	-4	-4	-1	-2	-1	-4	-1	4	-5	-3	-4	14	4	-3
	Y	-2	-2	-1	-3	0	-2	-3	-4	2	-1	0	-2	0	5	-3	-2	-2	4	8	-1
35	V	0	-2	-2	-3	0	-2	-2	-3	-2	3	2	-2	2	0	-2	-1	0	-3	-1	3

For purposes herein, a "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix reproduced herein above.

As is also well known in the art, relatedness of
5 nucleic acids can also be characterized using a functional test, the ability of the two nucleic acids to base-pair to one another at defined hybridization stringencies.

It is, therefore, another aspect of the invention to provide isolated nucleic acids not only identical in sequence
10 to those described with particularity herein, but also to provide isolated nucleic acids ("cross-hybridizing nucleic acids") that hybridize under high stringency conditions (as defined herein below) to all or to a portion of various of the isolated STTK nucleic acids of the present invention
15 ("reference nucleic acids"), as well as cross-hybridizing nucleic acids that hybridize under moderate stringency conditions to all or to a portion of various of the isolated STTK nucleic acids of the present invention.

Such cross-hybridizing nucleic acids are useful,
20 *inter alia*, as probes for, and to drive expression of, proteins related to the proteins of the present invention as alternative isoforms, homologues, paralogues, and orthologues. Particularly useful orthologues are those from other primate species, such as chimpanzee, rhesus macaque, monkey, baboon,
25 orangutan, and gorilla; from rodents, such as rats, mice, guinea pigs; from lagomorphs, such as rabbits; and from domestic livestock, such as cow, pig, sheep, horse, goat and chicken.

For purposes herein, high stringency conditions are
30 defined as aqueous hybridization (*i.e.*, free of formamide) in 6X SSC (where 20X SSC contains 3.0 M NaCl and 0.3 M sodium citrate), 1% SDS at 65°C for at least 8 hours, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. For purposes

herein, moderate stringency conditions are defined as aqueous hybridization (i.e., free of formamide) in 6X SSC, 1% SDS at 65°C for at least 8 hours, followed by one or more washes in 2x SSC, 0.1% SDS at room temperature.

5 The hybridizing portion of the reference nucleic acid is typically at least 15 nucleotides in length, often at least 17 nucleotides in length. Often, however, the hybridizing portion of the reference nucleic acid is at least 20 nucleotides in length, 25 nucleotides in length, and even 30
10 nucleotides, 35 nucleotides, 40 nucleotides, and 50 nucleotides in length. Of course, cross-hybridizing nucleic acids that hybridize to a larger portion of the reference nucleic acid - for example, to a portion of at least 50 nt, at least 100 nt, at least 150 nt, 200 nt, 250 nt, 300 nt, 350 nt, 400 nt, 450
15 nt, or 500 nt or more - or even to the entire length of the reference nucleic acid, are also useful.

 The hybridizing portion of the cross-hybridizing nucleic acid is at least 75% identical in sequence to at least a portion of the reference nucleic acid. Typically, the
20 hybridizing portion of the cross-hybridizing nucleic acid is at least 80%, often at least 85%, 86%, 87%, 88%, 89% or even at least 90% identical in sequence to at least a portion of the reference nucleic acid. Often, the hybridizing portion of the cross-hybridizing nucleic acid will be at least 91%, 92%, 93%,
25 94%, 95%, 96%, 97%, 98%, or 99% identical in sequence to at least a portion of the reference nucleic acid sequence. At times, the hybridizing portion of the cross-hybridizing nucleic acid will be at least 99.5% identical in sequence to at least a portion of the reference nucleic acid.

30 The invention also provides fragments of various of the isolated nucleic acids of the present invention.

 By "fragments" of a reference nucleic acid is here intended isolated nucleic acids, however obtained, that have a

nucleotide sequence identical to a portion of the reference nucleic acid sequence, which portion is at least 17 nucleotides and less than the entirety of the reference nucleic acid. As so defined, "fragments" need not be obtained by physical
5 fragmentation of the reference nucleic acid, although such provenance is not thereby precluded.

In theory, an oligonucleotide of 17 nucleotides is of sufficient length as to occur at random less frequently than once in the three gigabase human genome, and thus to provide a
10 nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. As is well known, further specificity can be obtained by probing nucleic acid samples of subgenomic complexity, and/or by using plural fragments as short as 17 nucleotides in length
15 collectively to prime amplification of nucleic acids, as, e.g., by polymerase chain reaction (PCR).

As further described herein below, nucleic acid fragments that encode at least 6 contiguous amino acids (*i.e.*, fragments of 18 nucleotides or more) are useful in directing
20 the expression or the synthesis of peptides that have utility in mapping the epitopes of the protein encoded by the reference nucleic acid. See, e.g., Geysen et al., "Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid," *Proc. Natl. Acad. Sci. USA* 81:3998-
25 4002 (1984); and U.S. Pat. Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties.

As further described herein below, fragments that encode at least 8 contiguous amino acids (*i.e.*, fragments of 24
30 nucleotides or more) are useful in directing the expression or the synthesis of peptides that have utility as immunogens. See, e.g., Lerner, "Tapping the immunological repertoire to produce antibodies of predetermined specificity," *Nature*

299:592-596 (1982); Shinnick et al., "Synthetic peptide immunogens as vaccines," *Annu. Rev. Microbiol.* 37:425-46 (1983); Sutcliffe et al., "Antibodies that react with predetermined sites on proteins," *Science* 219:660-6 (1983), the
5 disclosures of which are incorporated herein by reference in their entireties.

The nucleic acid fragment of the present invention is thus at least 17 nucleotides in length, typically at least 18 nucleotides in length, and often at least 24 nucleotides in
10 length. Often, the nucleic acid of the present invention is at least 25 nucleotides in length, and even 30 nucleotides, 35 nucleotides, 40 nucleotides, or 45 nucleotides in length. Of course, larger fragments having at least 50 nt, at least 100 nt, at least 150 nt, 200 nt, 250 nt, 300 nt, 350 nt, 400 nt,
15 450 nt, or 500 nt or more are also useful, and at times preferred.

Having been based upon the mining of genomic sequence, rather than upon surveillance of expressed message, the present invention further provides isolated genome-derived
20 nucleic acids that include portions of the STTK gene.

The invention particularly provides genome-derived single exon probes.

As further described in commonly owned and copending U.S. patent application serial nos. 09/864,761, filed May 23, 2001; 09/774,203, filed January 29, 2001; and 09/632,366, filed
25 August 3, 2000, the disclosures of which are incorporated herein by reference in their entireties, "a single exon probe" comprises at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to
30 transcript-derived nucleic acids that include the reference exon. The single exon probe will not, however, hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon and instead consist of one or more

exons that are found adjacent to the reference exon in the genome.

Genome-derived single exon probes typically further comprise, contiguous to a first end of the exon portion, a
5 first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. Often, the genome-derived single exon probe further comprises, contiguous to a second end of the exonic portion, a second intronic and/or intergenic sequence that is identically contiguous to the exon
10 in the genome.

The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids. Accordingly,
15 the exon portion is at least 17 nucleotides, typically at least 18 nucleotides, 20 nucleotides, 24 nucleotides, 25 nucleotides or even 30, 35, 40, 45, or 50 nucleotides in length, and can usefully include the entirety of the exon, up to 100 nt, 150 nt, 200 nt, 250 nt, 300 nt, 350 nt, 400 nt or even 500 nt or
20 more in length.

The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon, that is, be unable to hybridize detectably under high stringency conditions to
25 nucleic acids that lack the reference exon but include one or more exons that are found adjacent to the reference exon the genome.

Given variable spacing of exons through eukaryotic genomes, the maximum length of single exon probes of the
30 present invention is typically no more than 25 kb, often no more than 20 kb, 15 kb, 10 kb or 7.5 kb, or even no more than 5 kb, 4 kb, 3 kb, or even no more than about 2.5 kb in length.

The genome-derived single exon probes of the present invention can usefully include at least a first terminal priming sequence not found in contiguity with the rest of the probe sequence in the genome, and often will contain a second
5 terminal priming sequence not found in contiguity with the rest of the probe sequence in the genome.

The present invention also provides isolated genome-derived nucleic acids that include nucleic acid sequence elements that control transcription of the STTK gene.

10 With a complete draft of the human genome now available, genomic sequences that are within the vicinity of the STTK coding region (and that are additional to those described with particularity herein) can readily be obtained by PCR amplification.

15 The isolated nucleic acids of the present invention can be composed of natural nucleotides in native 5'-3' phosphodiester internucleoside linkage - e.g., DNA or RNA - or can contain any or all of nonnatural nucleotide analogues, nonnative internucleoside bonds, or post-synthesis
20 modifications, either throughout the length of the nucleic acid or localized to one or more portions thereof.

As is well known in the art, when the isolated nucleic acid is used as a hybridization probe, the range of such nonnatural analogues, nonnative internucleoside bonds, or
25 post-synthesis modifications will be limited to those that permit sequence-discriminating basepairing of the resulting nucleic acid. When used to direct expression or RNA or protein *in vitro* or *in vivo*, the range of such nonnatural analogues, nonnative internucleoside bonds, or post-synthesis
30 modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the

range of such changes will be limited to those that do not confer toxicity upon the isolated nucleic acid.

For example, when desired to be used as probes, the isolated nucleic acids of the present invention can usefully include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens.

Common radiolabeled analogues include those labeled with ^{33}P , ^{32}P , and ^{35}S , such as α - ^{32}P -dATP, α - ^{32}P -dCTP, α - ^{32}P -dGTP, α - ^{32}P -dTTP, α - ^{32}P -3'dATP, α - ^{32}P -ATP, α - ^{32}P -CTP, α - ^{32}P -GTP, α - ^{32}P -UTP, α - ^{35}S -dATP, γ - ^{35}S -GTP, γ - ^{33}P -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA).

Protocols are available for custom synthesis of nucleotides having other fluorophores. Henegariu et al., "Custom Fluorescent-Nucleotide Synthesis as an Alternative Method for Nucleic Acid Labeling," *Nature Biotechnol.* 18:345 -

348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, 5 Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, 10 USA).

As another example, when desired to be used for antisense inhibition of transcription or translation, the isolated nucleic acids of the present invention can usefully include altered, often nuclease-resistant, internucleoside 15 bonds. See Hartmann et al. (eds.), Manual of Antisense Methodology (Perspectives in Antisense Science), Kluwer Law International (1999) (ISBN:079238539X); Stein et al. (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (cover (1998) (ISBN: 0471172790); Chadwick et al. (eds.), 20 Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997) (ISBN: 0471972797), the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired also when the isolated nucleic acid of the present 25 invention is to be used for targeted gene correction, Gamper et al., *Nucl. Acids Res.* 28(21):4332-4339 (2000), the disclosures of which are incorporated herein by reference in its entirety.

Modified oligonucleotide backbones often preferred when the nucleic acid is to be used for antisense purposes are, 30 for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates,

phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative U.S. patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U.S. Pat. Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties.

Preferred modified oligonucleotide backbones for antisense use that do not include a phosphorus atom have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Pat. Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134;

5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938;
5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307;
5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240;
5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312;

5 5,633,360; 5,677,437; and 5,677,439, the disclosures of which
are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the
sugar and the internucleoside linkage are replaced with novel
groups, such as peptide nucleic acids (PNA).

10 In PNA compounds, the phosphodiester backbone of the
nucleic acid is replaced with an amide-containing backbone, in
particular by repeating N-(2-aminoethyl) glycine units linked
by amide bonds. Nucleobases are bound directly or indirectly
to aza nitrogen atoms of the amide portion of the backbone,
15 typically by methylene carbonyl linkages.

The uncharged nature of the PNA backbone provides
PNA/DNA and PNA/RNA duplexes with a higher thermal stability
than is found in DNA/DNA and DNA/RNA duplexes, resulting from
the lack of charge repulsion between the PNA and DNA or RNA
20 strand. In general, the T_m of a PNA/DNA or PNA/RNA duplex is
1°C higher per base pair than the T_m of the corresponding
DNA/DNA or DNA/RNA duplex (in 100 mM NaCl).

The neutral backbone also allows PNA to form stable
DNA duplexes largely independent of salt concentration. At low
25 ionic strength, PNA can be hybridized to a target sequence at
temperatures that make DNA hybridization problematic or
impossible. And unlike DNA/DNA duplex formation, PNA
hybridization is possible in the absence of magnesium.
Adjusting the ionic strength, therefore, is useful if competing
30 DNA or RNA is present in the sample, or if the nucleic acid
being probed contains a high level of secondary structure.

PNA also demonstrates greater specificity in binding
to complementary DNA. A PNA/DNA mismatch is more destabilizing

than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by 8-20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by 4-16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater.

Additionally, nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. As a result, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro*. In addition, PNA is stable over a wide pH range.

Because its backbone is formed from amide bonds, PNA can be synthesized using a modified peptide synthesis protocol.

PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Pat. Nos. 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference; automated PNA synthesis is readily achievable on commercial synthesizers (see, e.g., "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA chemistry and applications are reviewed, *inter alia*, in Ray et al., *FASEB J.* 14(9):1041-60 (2000); Nielsen et al., *Pharmacol Toxicol.* 86(1):3-7 (2000); Larsen et al., *Biochim Biophys Acta.* 1489(1):159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3):353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1):71-5 (1999), the disclosures of which are incorporated herein by reference in their entirety.

Differences from nucleic acid compositions found in nature - e.g., nonnative bases, altered internucleoside linkages, post-synthesis modification - can be present

throughout the length of the nucleic acid or can, instead, usefully be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and demonstrated utility for targeted gene repair, as further described in U.S. Pat. Nos. 5,760,012 and 5,731,181, the disclosures of which are incorporated herein by reference in their entireties. As another example, chimeric nucleic acids comprising both DNA and PNA have been demonstrated to have utility in modified PCR reactions. See Misra et al., *Biochem.* 37: 1917-1925 (1998); see also Finn et al., *Nucl. Acids Res.* 24: 3357-3363 (1996), incorporated herein by reference.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér et al., *Curr. Opin. Biotechnol.* 12:11-15 (2001); Escude et al., *Proc. Natl. Acad. Sci. USA* 14;96(19):10603-7 (1999); Nilsson et al., *Science* 265(5181):2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth et al., *Biochim. Biophys. Acta.* 1489(1):181-206 (1999); Fox, *Curr. Med. Chem.* 7(1):17-37 (2000); Kochetkova et al., *Methods Mol. Biol.* 130:189-201 (2000); Chan et al., *J. Mol. Med.* 75(4):267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

The nucleic acids of the present invention can be detectably labeled.

Commonly-used labels include radionuclides, such as ^{32}P , ^{33}P , ^{35}S , ^3H (and for NMR detection, ^{13}C and ^{15}N), haptens that can be detected by specific antibody or high affinity binding partner (such as avidin), and fluorophores.

5 As noted above, detectable labels can be incorporated by inclusion of labeled nucleotide analogues in the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing,
10 and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, e.g., from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach.

 Analogues can also be incorporated during automated
15 solid phase chemical synthesis.

 As is well known, labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

20 Various other post-synthetic approaches permit internal labeling of nucleic acids.

 For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA
25 to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); see Alers et al., *Genes, Chromosomes & Cancer*, Vol. 25, pp. 301 - 305
30 (1999); Jelsma et al., *J. NIH Res.* 5:82 (1994); Van Belkum et al., *BioTechniques* 16:148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTagTM Reagent, Vector

Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or
5 other marker.

Multiple independent or interacting labels can be incorporated into the nucleic acids of the present invention.

For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included
10 to report specific hybridization through release of fluorescence quenching, Tyagi et al., *Nature Biotechnol.* 14: 303-308 (1996); Tyagi et al., *Nature Biotechnol.* 16, 49-53 (1998); Sokol et al., *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis et al., *Science* 279:1228-1229
15 (1998); Marras et al., *Genet. Anal.* 14: 151-156 (1999); U.S. Pat. Nos. 5,846,726, 5,925,517, 5,925,517, or to report exonucleotidic excision, U.S. Pat. No. 5,538,848; Holland et al., *Proc. Natl. Acad. Sci. USA* 88:7276-7280 (1991); Heid et al., *Genome Res.* 6(10):986-94 (1996); Kuimelis et al., *Nucleic
20 Acids Symp Ser.* (37):255-6 (1997); U.S. Patent No. 5,723,591, the disclosures of which are incorporated herein by reference in their entireties.

So labeled, the isolated nucleic acids of the present invention can be used as probes, as further described below.

25 Nucleic acids of the present invention can also usefully be bound to a substrate. The substrate can porous or solid, planar or non-planar, unitary or distributed; the bond can be covalent or noncovalent. Bound to a substrate, nucleic acids of the present invention can be used as probes in their
30 unlabeled state.

For example, the nucleic acids of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, nylon, or

positively-charged derivatized nylon; so attached, the nucleic acids of the present invention can be used to detect STTK nucleic acids present within a labeled nucleic acid sample, either a sample of genomic nucleic acids or a sample of transcript-derived nucleic acids, e.g. by reverse dot blot.

The nucleic acids of the present invention can also usefully be bound to a solid substrate, such as glass, although other solid materials, such as amorphous silicon, crystalline silicon, or plastics, can also be used. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof.

Typically, the solid substrate will be rectangular, although other shapes, particularly disks and even spheres, present certain advantages. Particularly advantageous alternatives to glass slides as support substrates for array of nucleic acids are optical discs, as described in Demers, "Spatially Addressable Combinatorial Chemical Arrays in CD-ROM Format," international patent publication WO 98/12559, incorporated herein by reference in its entirety.

The nucleic acids of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof.

The nucleic acids of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, e.g. on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at

higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

The isolated nucleic acids of the present invention can be used as hybridization probes to detect, characterize, and quantify STTK nucleic acids in, and isolate STTK nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

For example, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the STTK genomic locus, such as deletions, insertions, translocations, and duplications of the STTK genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. See, e.g., Andreeff et al. (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999) (ISBN: 0471013455), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, e.g., Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acids of the present invention can be used as probes to isolate genomic clones that include the nucleic acids of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

The isolated nucleic acids of the present invention can also be used as probes to detect, characterize, and quantify STTK nucleic acids in, and isolate STTK nucleic acids from, transcript-derived nucleic acid samples.

5 For example, the isolated nucleic acids of the present invention can be used as hybridization probes to detect, characterize by length, and quantify STTK mRNA by northern blot of total or poly-A⁺- selected RNA samples. For example, the isolated nucleic acids of the present invention
10 can be used as hybridization probes to detect, characterize by location, and quantify STTK message by *in situ* hybridization to tissue sections (see, e.g., Schwarczacher et al., In Situ Hybridization, Springer-Verlag New York (2000) (ISBN: 0387915966), the disclosure of which is incorporated herein by
15 reference in its entirety). For example, the isolated nucleic acids of the present invention can be used as hybridization probes to measure the representation of STTK clones in a cDNA library. For example, the isolated nucleic acids of the present invention can be used as hybridization probes to
20 isolate STTK nucleic acids from cDNA libraries, permitting sequence level characterization of STTK messages, including identification of deletions, insertions, truncations – including deletions, insertions, and truncations of exons in alternatively spliced forms – and single nucleotide
25 polymorphisms.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook et al., Molecular Cloning: A Laboratory Manual (3rd ed.), Cold Spring Harbor
30 Laboratory Press (2001) (ISBN: 0879695773); Ausubel et al. (eds.), Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology (4th ed.), John Wiley & Sons, 1999 (ISBN: 047132938X); and Walker et al.

(eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000) (ISBN: 0896034593), the disclosures of which are incorporated herein by reference in their entirety.

As described in the Examples herein below, the
5 nucleic acids of the present invention can also be used to detect and quantify STTK nucleic acids in transcript-derived samples - that is, to measure expression of the STTK gene - when included in a microarray. Measurement of STTK expression has particular utility in the therapy and diagnosis of cancer,
10 as further described in the Examples herein below.

As would be readily apparent to one of skill in the art, each STTK nucleic acid probe - whether labeled, substrate-bound, or both - is thus currently available for use as a tool for measuring the level of STTK expression in each of the
15 tissues in which expression has already been confirmed, notably adult liver, bone marrow, brain, colon, fetal liver, heart, kidney, lung, placenta, and skeletal muscle as well as a cell line HeLa. The utility is specific to the probe: under high stringency conditions, the probe reports the level of
20 expression of message specifically containing that portion of the STTK gene included within the probe.

Measuring tools are well known in many arts, not just in molecular biology, and are known to possess credible, specific, and substantial utility. For example, U.S. Patent
25 No. 6,016,191 describes and claims a tool for measuring characteristics of fluid flow in a hydrocarbon well; U.S. Patent No. 6,042,549 describes and claims a device for measuring exercise intensity; U.S. Patent No. 5,889,351 describes and claims a device for measuring viscosity and for
30 measuring characteristics of a fluid; U.S. Patent No. 5,570,694 describes and claims a device for measuring blood pressure; U.S. Patent No. 5,930,143 describes and claims a device for measuring the dimensions of machine tools; U.S. Patent No.

5,279,044 describes and claims a measuring device for determining an absolute position of a movable element; U.S. Patent No. 5,186,042 describes and claims a device for measuring action force of a wheel; and U.S. Patent No.

5 4,246,774 describes and claims a device for measuring the draft of smoking articles such as cigarettes.

As for tissues not yet demonstrated to express STTK, the STTK nucleic acid probes of the present invention are currently available as tools for surveying such tissues to
10 detect the presence of STTK nucleic acids.

Survey tools - *i.e.*, tools for determining the presence and/or location of a desired object by search of an area - are well known in many arts, not just in molecular biology, and are known to possess credible, specific, and
15 substantial utility. For example, U.S. Patent No. 6,046,800 describes and claims a device for surveying an area for objects that move; U.S. Patent No. 6,025,201 describes and claims an apparatus for locating and discriminating platelets from non-platelet particles or cells on a cell-by-cell basis in a
20 whole blood sample; U.S. Patent No. 5,990,689 describes and claims a device for detecting and locating anomalies in the electromagnetic protection of a system; U.S. Patent No. 5,984,175 describes and claims a device for detecting and identifying wearable user identification units; U.S. Patent No.
25 3,980,986 ("Oil well survey tool"), describes and claims a tool for finding the position of a drill bit working at the bottom of a borehole.

As noted above, the nucleic acid probes of the present invention are useful in constructing microarrays; the
30 microarrays, in turn, are products of manufacture that are useful for measuring and for surveying gene expression.

When included on a microarray, each STTK nucleic acid probe makes the microarray specifically useful for detecting

that portion of the STTK gene included within the probe, thus imparting upon the microarray device the ability to detect a signal where, absent such probe, it would have reported no signal. This utility makes each individual probe on such
5 microarray akin to an antenna, circuit, firmware or software element included in an electronic apparatus, where the antenna, circuit, firmware or software element imparts upon the apparatus the ability newly and additionally to detect signal in a portion of the radio-frequency spectrum where previously
10 it could not; such devices are known to have specific, substantial, and credible utility.

Changes in the level of expression need not be observed for the measurement of expression to have utility.

For example, where gene expression analysis is used
15 to assess toxicity of chemical agents on cells, the failure of the agent to change a gene's expression level is evidence that the drug likely does not affect the pathway of which the gene's expressed protein is a part. Analogously, where gene expression analysis is used to assess side effects of
20 pharmacologic agents - whether in lead compound discovery or in subsequent screening of lead compound derivatives - the inability of the agent to alter a gene's expression level is evidence that the drug does not affect the pathway of which the gene's expressed protein is a part.

25 WO 99/58720, incorporated herein by reference in its entirety, provides methods for quantifying the relatedness of a first and second gene expression profile and for ordering the relatedness of a plurality of gene expression profiles, without regard to the identity or function of the genes whose
30 expression is used in the calculation.

Gene expression analysis, including gene expression analysis by microarray hybridization, is, of course, principally a laboratory-based art. Devices and apparatus used

principally in laboratories to facilitate laboratory research are well-established to possess specific, substantial, and credible utility. For example, U.S. Patent No. 6,001,233 describes and claims a gel electrophoresis apparatus having a cam-activated clamp; for example, U.S. Patent No. 6,051,831 describes and claims a high mass detector for use in time-of-flight mass spectrometers; for example, U.S. Patent NO. 5,824,269 describes and claims a flow cytometer—as is well known, few gel electrophoresis apparatuses, TOF-MS devices, or flow cytometers are sold for consumer use.

Indeed, and in particular, nucleic acid microarrays, as devices intended for laboratory use in measuring gene expression, are well-established to have specific, substantial and credible utility. Thus, the microarrays of the present invention have at least the specific, substantial and credible utilities of the microarrays claimed as devices and articles of manufacture in the following U.S. patents, the disclosures of each of which is incorporated herein by reference: U.S. Patent Nos. 5,445,934 ("Array of oligonucleotides on a solid substrate"); 5,744,305 ("Arrays of materials attached to a substrate"); and 6,004,752 ("Solid support with attached molecules").

Genome-derived single exon probes and genome-derived single exon probe microarrays have the additional utility, *inter alia*, of permitting high-throughput detection of splice variants of the nucleic acids of the present invention, as further described in copending and commonly owned U.S. Patent application no. 09/632,366, filed August 3, 2000, the disclosure of which is incorporated herein by reference in its entirety.

The isolated nucleic acids of the present invention can also be used to prime synthesis of nucleic acid, for

purpose of either analysis or isolation, using mRNA, cDNA, or genomic DNA as template.

For use as primers, at least 17 contiguous nucleotides of the isolated nucleic acids of the present invention will be used. Often, at least 18, 19, or 20 contiguous nucleotides of the nucleic acids of the present invention will be used, and on occasion at least 20, 22, 24, or 25 contiguous nucleotides of the nucleic acids of the present invention will be used, and even 30 nucleotides or more of the nucleic acids of the present invention can be used to prime specific synthesis.

The nucleic acid primers of the present invention can be used, for example, to prime first strand cDNA synthesis on an mRNA template.

Such primer extension can be done directly to analyze the message. Alternatively, synthesis on an mRNA template can be done to produce first strand cDNA. The first strand cDNA can thereafter be used, *inter alia*, directly as a single-stranded probe, as above-described, as a template for sequencing - permitting identification of alterations, including deletions, insertions, and substitutions, both normal allelic variants and mutations associated with abnormal phenotypes- or as a template, either for second strand cDNA synthesis (e.g., as an antecedent to insertion into a cloning or expression vector), or for amplification.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (see, e.g., U.S. Pat. No. 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

As another example, the nucleic acid primers of the present invention can be used to prime amplification of STTK

nucleic acids, using transcript-derived or genomic DNA as template.

Primer-directed amplification methods are now well-established in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR (Basics: From Background to Bench), Springer Verlag (2000) (ISBN: 0387916008); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999) (ISBN: 0123721857); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998) (ISBN: 0123721822); Newton *et al.*, PCR, Springer-Verlag New York (1997) (ISBN: 0387915060); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996) (ISBN: 047195697X); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996) (ISBN: 0896033430); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995) (ISBN: 0199634254), the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998 (ISBN: 1881299147); Siebert (ed.), PCR Technique: RT-PCR, Eaton Publishing Company/BioTechniques Books (1995) (ISBN: 1881299139), the disclosure of which is incorporated herein by reference in its entirety.

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1):21-7 (2001); U.S. Patent Nos. 6,235,502, 6,221,603, 6,210,884, 6,183,960, 5,854,033, 5,714,320, 5,648,245, and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other

techniques to facilitate SNP detection. See, e.g., Lizardi et al., *Nature Genet.* 19(3):225-32 (1998).

As further described below, nucleic acids of the present invention, inserted into vectors that flank the nucleic acid insert with a phage promoter, such as T7, T3, or SP6 promoter, can be used to drive *in vitro* expression of RNA complementary to either strand of the nucleic acid of the present invention. The RNA can be used, *inter alia*, as a single-stranded probe, in cDNA-mRNA subtraction, or for *in vitro* translation.

As will be further discussed herein below, nucleic acids of the present invention that encode STTK protein or portions thereof can be used, *inter alia*, to express the STTK proteins or protein fragments, either alone, or as part of fusion proteins.

Expression can be from genomic nucleic acids of the present invention, or from transcript-derived nucleic acids of the present invention.

Where protein expression is effected from genomic DNA, expression will typically be effected in eukaryotic, typically mammalian, cells capable of splicing introns from the initial RNA transcript. Expression can be driven from episomal vectors, such as EBV-based vectors, or can be effected from genomic DNA integrated into a host cell chromosome. As will be more fully described below, where expression is from transcript-derived (or otherwise intron-less) nucleic acids of the present invention, expression can be effected in wide variety of prokaryotic or eukaryotic cells.

Expressed *in vitro*, the protein, protein fragment, or protein fusion can thereafter be isolated, to be used, *inter alia*, as a standard in immunoassays specific for the proteins, or protein isoforms, of the present invention; to be used as a therapeutic agent, e.g., to be administered as passive

replacement therapy in individuals deficient in the proteins of the present invention, or to be administered as a vaccine; to be used for *in vitro* production of specific antibody, the antibody thereafter to be used, e.g., as an analytical reagent for detection and quantitation of the proteins of the present invention or to be used as an immunotherapeutic agent.

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the proteins of the present invention. *In vivo* expression can be driven from a vector - typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV) - for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Pat. Nos. 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; 6,204,250, the disclosures of which are incorporated herein by reference in their entireties.

The nucleic acids of the present invention can also be used for antisense inhibition of transcription or translation. See Phillips (ed.), Antisense Technology, Part B, Methods in Enzymology Vol. 314, Academic Press, Inc. (1999) (ISBN: 012182215X); Phillips (ed.), Antisense Technology, Part A, Methods in Enzymology Vol. 313, Academic Press, Inc. (1999) (ISBN: 0121822141); Hartmann et al. (eds.), Manual of Antisense Methodology (Perspectives in Antisense Science), Kluwer Law International (1999) (ISBN:079238539X); Stein et al. (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (cover (1998) (ISBN: 0471172790); Agrawal et al. (eds.), Antisense Research and Application, Springer-Verlag New York, Inc. (1998) (ISBN: 3540638334); Lichtenstein et al. (eds.), Antisense

Technology: A Practical Approach, Vol. 185, Oxford University Press, INC. (1998) (ISBN: 0199635838); Gibson (ed.), Antisense and Ribozyme Methodology: Laboratory Companion, Chapman & Hall (1997) (ISBN: 3826100794); Chadwick et al. (eds.),

- 5 Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997) (ISBN: 0471972797), the disclosures of which are incorporated herein by reference in their entireties.

10 Nucleic acids of the present invention, particularly cDNAs of the present invention, that encode full-length human STTK protein isoforms, have additional, well-recognized, immediate, real world utility as commercial products of manufacture suitable for sale.

15 For example, Invitrogen Corp. (Carlsbad, CA, USA), through its Research Genetics subsidiary, sells full length human cDNAs cloned into one of a selection of expression vectors as GeneStorm® expression-ready clones; utility is specific for the gene, since each gene is capable of being ordered separately and has a distinct catalogue number, and
20 utility is substantial, each clone selling for \$650.00 US. Similarly, Incyte Genomics (Palo Alto, CA, USA) sells clones from public and proprietary sources in multi-well plates or individual tubes.

25 Nucleic acids of the present invention that include genomic regions encoding the STTK protein, or portions thereof, have yet further utilities.

30 For example, genomic nucleic acids of the present invention can be used as amplification substrates, e.g. for preparation of genome-derived single exon probes of the present invention, as described above and in copending and commonly-owned U.S. patent application nos. 09/864,761, filed May 23, 2001, 09/774,203, filed January 29, 2001, and 09/632,366, filed

August 3, 2000, the disclosures of which are incorporated herein by reference in their entireties.

As another example, genomic nucleic acids of the present invention can be integrated non-homologously into the genome of somatic cells, e.g. CHO cells, COS cells, or 293 cells, with or without amplification of the insertional locus, in order, e.g., to create stable cell lines capable of producing the proteins of the present invention.

As another example, more fully described herein below, genomic nucleic acids of the present invention can be integrated nonhomologously into embryonic stem (ES) cells to create transgenic non-human animals capable of producing the proteins of the present invention.

Genomic nucleic acids of the present invention can also be used to target homologous recombination to the STTK locus. See, e.g., U.S. Patent Nos. 6,187,305; 6,204,061; 5,631,153; 5,627,059; 5,487,992; 5,464,764; 5,614,396; 5,527,695 and 6,063,630; and Kmiec et al. (eds.), Gene Targeting Protocols, Vol. 133, Humana Press (2000) (ISBN: 0896033600); Joyner (ed.), Gene Targeting: A Practical Approach, Oxford University Press, Inc. (2000) (ISBN: 0199637938); Sedivy et al., Gene Targeting, Oxford University Press (1998) (ISBN: 071677013X); Tymms et al. (eds.), Gene Knockout Protocols, Humana Press (2000) (ISBN: 0896035727); Mak et al. (eds.), The Gene Knockout FactsBook, Vol. 2, Academic Press, Inc. (1998) (ISBN: 0124660444); Torres et al., Laboratory Protocols for Conditional Gene Targeting, Oxford University Press (1997) (ISBN: 019963677X); Vega (ed.), Gene Targeting, CRC Press, LLC (1994) (ISBN: 084938950X), the disclosures of which are incorporated herein by reference in their entireties.

Where the genomic region includes transcription regulatory elements, homologous recombination can be used to

alter the expression of STTK, both for purpose of *in vitro* production of STTK protein from human cells, and for purpose of gene therapy. See, e.g., U.S. Pat. Nos. 5,981,214, 6,048,524; 5,272,071.

5 Fragments of the nucleic acids of the present invention smaller than those typically used for homologous recombination can also be used for targeted gene correction or alteration, possibly by cellular mechanisms different from those engaged during homologous recombination.

10 For example, partially duplexed RNA/DNA chimeras have been shown to have utility in targeted gene correction, U.S. Pat. Nos. 5,945,339, 5,888,983, 5,871,984, 5,795,972, 5,780,296, 5,760,012, 5,756,325, 5,731,181, the disclosures of which are incorporated herein by reference in their entireties.

15 So too have small oligonucleotides fused to triplexing domains have been shown to have utility in targeted gene correction, Culver et al., "Correction of chromosomal point mutations in human cells with bifunctional oligonucleotides," *Nature Biotechnol.* 17(10):989-93 (1999), as have oligonucleotides
20 having modified terminal bases or modified terminal internucleoside bonds, Gamper et al., *Nucl. Acids Res.* 28(21):4332-9 (2000), the disclosures of which are incorporated herein by reference.

 The isolated nucleic acids of the present invention
25 can also be used to provide the initial substrate for recombinant engineering of STTK protein variants having desired phenotypic improvements. Such engineering includes, for example, site-directed mutagenesis, random mutagenesis with subsequent functional screening, and more elegant schemes for
30 recombinant evolution of proteins, as are described, *inter alia*, in U.S. Pat. Nos. 6,180,406; 6,165,793; 6,117,679; and 6,096,548, the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acids of the present invention can be obtained by using the labeled probes of the present invention to probe nucleic acid samples, such as genomic libraries, cDNA libraries, and mRNA samples, by standard techniques. Nucleic acids of the present invention can also be obtained by amplification, using the nucleic acid primers of the present invention, as further demonstrated in Example 1, herein below.

Nucleic acids of the present invention of fewer than about 100 nt can also be synthesized chemically, typically by solid phase synthesis using commercially available automated synthesizers.

"Full Length" Human STTK Nucleic Acids

In a first series of nucleic acid embodiments, the invention provides isolated nucleic acids that encode the entirety of the STTK protein. As discussed above, the "full-length" nucleic acids of the present invention can be used, *inter alia*, to express full length STTK protein. The full-length nucleic acids can also be used as nucleic acid probes; used as probes, the isolated nucleic acids of these embodiments will hybridize to STTK.

In a first such embodiment, the invention provides an isolated nucleic acid comprising (i) the nucleotide sequence of SEQ ID NO: 1, or (ii) the complement of (i). SEQ ID NO: 1 presents the entire cDNA of human STTK, including the 3' untranslated (UT) region.

In a second embodiment, the invention provides an isolated nucleic acid comprising (i) the nucleotide sequence of SEQ ID NO: 2, (ii) a degenerate variant of the nucleotide sequence of SEQ ID NO: 2, or (iii) the complement of (i) or (ii). SEQ ID NO: 2 presents the open reading frame (ORF) from SEQ ID NO: 1.

In a third embodiment, the invention provides an isolated nucleic acid comprising (i) a nucleotide sequence that encodes a polypeptide with the amino acid sequence of SEQ ID NO: 3 or (ii) the complement of a nucleotide sequence that encodes a polypeptide with the amino acid sequence of SEQ ID NO: 3. SEQ ID NO: 3 provides the amino acid sequence of human STTK.

In a fourth embodiment, the invention provides an isolated nucleic acid having a nucleotide sequence that (i) encodes a polypeptide having the sequence of SEQ ID NO: 3, (ii) encodes a polypeptide having the sequence of SEQ ID NO: 3 with conservative amino acid substitutions, or (iii) that is the complement of (i) or (ii), where SEQ ID NO: 3 provides the amino acid sequence of human STTK.

Selected Partial Nucleic Acids

In a second series of nucleic acid embodiments, the invention provides isolated nucleic acids that encode select portions of STTK. As will be further discussed herein below, these "partial" nucleic acids can be used, *inter alia*, to express specific portions of the STTK. These "partial" nucleic acids can also be used, *inter alia*, as nucleic probes.

In a first such embodiment, the invention provides an isolated nucleic acid comprising (i) a nucleotide sequence that encodes SEQ ID NO: 31 or (ii) the complement of a nucleotide sequence that encodes SEQ ID NO: 31, wherein the isolated nucleic acid is no more than about 100 kb in length, typically no more than about 75 kb in length, frequently no more than about 50 kb in length. SEQ ID NO: 31 is the amino acid sequence encoded by the portion of STTK not found in any earlier GenBank entries. Often, the isolated nucleic acids of this embodiment are no more than about 25 kb in length, often no more than

about 15 kb in length, and frequently no more than about 10 kb in length.

In another embodiment, the invention provides an isolated nucleic acid comprising (i) a nucleotide sequence that encodes SEQ ID NO: 31, (ii) a nucleotide sequence that encodes SEQ ID NO: 31 with conservative substitutions, or (iii) the complement of (i) or (ii), wherein the isolated nucleic acid is no more than about 100 kb in length, typically no more than about 75 kb in length, and often no more than about 50 kb in length. Often, the isolated nucleic acids of this embodiment are no more than about 25 kb in length, often no more than about 15 kb in length, and frequently no more than about 10 kb in length.

Cross-Hybridizing Nucleic Acids

In another nucleic acid embodiment, the invention provides isolated nucleic acids that hybridize to various of the STTK nucleic acids of the present invention. These cross-hybridizing nucleic acids can be used, *inter alia*, as probes for, and to drive expression of, proteins that are related to STTK of the present invention as further isoforms, homologues, paralogues, or orthologues.

In this embodiment, the invention provides an isolated nucleic acid comprising a sequence that hybridizes under high stringency conditions to a hybridization probe the nucleotide sequence of which (i) encodes a polypeptide having the sequence of SEQ ID NO: 31, (ii) encodes a polypeptide having the sequence of SEQ ID NO: 31 with conservative amino acid substitutions, or (iii) is the complement of (i) or (ii), wherein the isolated nucleic acid is no more than about 100 kb in length, typically no more than about 75 kb in length, and often no more than about 50 kb in length. Often, the isolated

nucleic acids of this embodiment are no more than about 25 kb in length, often no more than about 15 kb in length, and frequently no more than about 10 kb in length.

5 Particularly Useful Nucleic Acids

Particularly useful among the above-described nucleic acids are those that are expressed, or the complement of which are expressed, in adult liver, bone marrow, brain, colon, fetal
10 liver, heart, kidney, lung, placenta, and skeletal muscle as well as a cell line HeLa.

Also particularly useful among the above-described nucleic acids are those that encode, or the complement of which encode, a polypeptide having protein kinase activity.

15 Other particularly useful embodiments of the nucleic acids above-described are those that encode, or the complement of which encode, a polypeptide having a protein kinase domain.

Nucleic Acid Fragments

20 In another series of nucleic acid embodiments, the invention provides fragments of various of the isolated nucleic acids of the present invention which prove useful, *inter alia*, as nucleic acid probes, as amplification primers, and to direct
25 expression or synthesis of epitopic or immunogenic protein fragments.

 In a first such embodiment, the invention provides an isolated nucleic acid comprising (i) a nucleotide sequence that encodes a peptide of at least 8 contiguous amino acids of SEQ
30 ID NO: 31, (ii) a nucleotide sequence that encodes a peptide of at least 15 contiguous amino acids of SEQ ID NO: 31, or (iii) the complement of (i) or (ii), wherein the isolated nucleic acid is no more than about 100 kb in length, typically no more than about 75 kb in length, more typically no more than about

50 kb in length. Often, the isolated nucleic acids of this embodiment are no more than about 25 kb in length, often no more than about 15 kb in length, and frequently no more than about 10 kb in length.

5 In another embodiment, the invention provides an isolated nucleic acid comprising a nucleotide sequence that encodes (i) a polypeptide having the sequence of at least 8 contiguous amino acids of SEQ ID NO: 31 with conservative amino acid substitutions, (ii) a polypeptide having the sequence of
10 at least 15 contiguous amino acids of SEQ ID NO: 31 with conservative amino acid substitutions, (iii) a polypeptide having the sequence of at least 8 contiguous amino acids of SEQ ID NO: 31 with moderately conservative substitutions, (iv) a polypeptide having the sequence of at least 15 contiguous amino
15 acids of SEQ ID NO: 31 with moderately conservative substitutions, or (v) the complement of any of (i) - (iv), wherein the isolated nucleic acid is no more than about 100 kb in length, typically no more than about 75 kb in length, more typically no more than about 50 kb in length. Often, the
20 isolated nucleic acids of this embodiment are no more than about 25 kb in length, often no more than about 15 kb in length, and frequently no more than about 10 kb in length.

Single Exon Probes

25 The invention further provides genome-derived single exon probes having portions of no more than one exon of the STTK gene. As further described in commonly owned and copending U.S. patent application serial no. 09/632,366, filed
30 August 3, 2000 ("Methods and Apparatus for High Throughput Detection and Characterization of alternatively Spliced Genes"), the disclosure of which is incorporated herein by reference in its entirety, such single exon probes have

particular utility in identifying and characterizing splice variants. In particular, such single exon probes are useful for identifying and discriminating the expression of distinct isoforms of STTK.

5 In a first embodiment, the invention provides an isolated nucleic acid comprising a nucleotide sequence of no more than one portion of SEQ ID NOs: 4-16 or the complement of SEQ ID NOs: 4-16, wherein the portion comprises at least 17 contiguous nucleotides, 18 contiguous nucleotides, 20
10 contiguous nucleotides, 24 contiguous nucleotides, 25 contiguous nucleotides, or 50 contiguous nucleotides of any one of SEQ ID NOs: 4-16, or their complement. In a further embodiment, the exonic portion comprises the entirety of the referenced SEQ ID NO: or its complement.

15 In other embodiments, the invention provides isolated single exon probes having the nucleotide sequence of any one of SEQ ID NOs: 17-29.

Transcription Control Nucleic Acids

20 In another aspect, the present invention provides genome-derived isolated nucleic acids that include nucleic acid sequence elements that control transcription of the STTK gene.

These nucleic acids can be used, *inter alia*, to drive
25 expression of heterologous coding regions in recombinant constructs, thus conferring upon such heterologous coding regions the expression pattern of the native STTK gene. These nucleic acids can also be used, conversely, to target heterologous transcription control elements to the STTK genomic
30 locus, altering the expression pattern of the STTK gene itself.

In a first such embodiment, the invention provides an isolated nucleic acid comprising the nucleotide sequence of SEQ ID NO: 30 or its complement, wherein the isolated nucleic acid

is no more than about 100 kb in length, typically no more than about 75 kb in length, more typically no more than about 50 kb in length. Often, the isolated nucleic acids of this embodiment are no more than about 25 kb in length, often no more than about 15 kb in length, and frequently no more than about 10 kb in length.

In another embodiment, the invention provides an isolated nucleic acid comprising at least 17, 18, 20, 24, or 25 nucleotides of the sequence of SEQ ID NO: 30 or its complement, wherein the isolated nucleic acid is no more than about 100 kb in length, typically no more than about 75 kb in length, more typically no more than about 50 kb in length. Often, the isolated nucleic acids of this embodiment are no more than about 25 kb in length, often no more than about 15 kb in length, and frequently no more than about 10 kb in length.

VECTORS AND HOST CELLS

In another aspect, the present invention provides vectors that comprise one or more of the isolated nucleic acids of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to

heterologous polypeptides. Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones et al. (eds.), Vectors: Cloning Applications : Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd 1998 (ISBN: 047196266X); Jones et al. (eds.), Vectors: Expression Systems : Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd, 1998 (ISBN:0471962678); Gacesa et al., Vectors: Essential Data, John Wiley & Sons, 1995 (ISBN: 0471948411); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co., 2000 (ISBN: 188129935X); Sambrook et al., Molecular Cloning: A Laboratory Manual (3rd ed.), Cold Spring Harbor Laboratory Press, 2001 (ISBN: 0879695773); Ausubel et al. (eds.), Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology (4th ed.), John Wiley & Sons, 1999 (ISBN: 047132938X), the disclosures of which are incorporated herein by reference in their entireties.

Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Typically, vectors are derived from virus, plasmid, prokaryotic or eukaryotic chromosomal elements, or some combination thereof, and include at least one origin of replication, at least one site for insertion of heterologous nucleic acid, typically in the form of a polylinker with multiple, tightly clustered, single cutting restriction sites, and at least one selectable marker, although some integrative vectors will lack an origin that is functional in the host to be chromosomally modified, and some vectors will lack selectable markers. Vectors of the present invention will

further include at least one nucleic acid of the present invention inserted into the vector in at least one location.

Where present, the origin of replication and selectable markers are chosen based upon the desired host cell or host cells; the host cells, in turn, are selected based upon the desired application.

For example, prokaryotic cells, typically *E. coli*, are typically chosen for cloning. In such case, vector replication is predicated on the replication strategies of coliform-infecting phage - such as phage lambda, M13, T7, T3 and P1 - or on the replication origin of autonomously replicating episomes, notably the ColE1 plasmid and later derivatives, including pBR322 and the pUC series plasmids. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: e.g., typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin, zeocin; auxotrophic markers can also be used.

As another example, yeast cells, typically *S. cerevisiae*, are chosen, *inter alia*, for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and to the ready ability to complement genetic defects using recombinantly expressed proteins, for identification of interacting protein components, e.g. through use of a two-hybrid system, and for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast.

Integrative YIp vectors do not replicate autonomously, but integrate, typically in single copy, into the yeast genome at low frequencies and thus replicate as part of the host cell chromosome; these vectors lack an origin of

replication that is functional in yeast, although they typically have at least one origin of replication suitable for propagation of the vector in bacterial cells. YE_p vectors, in contrast, replicate episomally and autonomously due to presence of the yeast 2 micron plasmid origin (2 μ m ori). The YC_p yeast centromere plasmid vectors are autonomously replicating vectors containing centromere sequences, CEN, and autonomously replicating sequences, ARS; the ARS sequences are believed to correspond to the natural replication origins of yeast chromosomes. YACs are based on yeast linear plasmids, denoted YL_p, containing homologous or heterologous DNA sequences that function as telomeres (TEL) *in vivo*, as well as containing yeast ARS (origins of replication) and CEN (centromeres) segments.

Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*. The URA3 and LYS2 yeast genes further permit negative selection based on specific inhibitors, 5-fluoro-orotic acid (FOA) and α -aminoadipic acid (α AA), respectively, that prevent growth of the prototrophic strains but allows growth of the *ura3* and *lys2* mutants, respectively. Other selectable markers confer resistance to, e.g., zeocin.

As yet another example, insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda* - e.g., Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA) - the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that

flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following cotransfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

As yet another example, mammalian cells are often chosen for expression of proteins intended as pharmaceutical agents, and are also chosen as host cells for screening of potential agonist and antagonists of a protein or a physiological pathway.

Where mammalian cells are chosen as host cells, vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, e.g., in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy.

Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (e.g.,

cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

For propagation of nucleic acids of the present invention that are larger than can readily be accommodated in
5 vectors derived from plasmids or virus, the invention further provides artificial chromosomes - BACs, YACs, PACs, and HACs - that comprise human STTK nucleic acids, often genomic nucleic acids.

The BAC system is based on the well-characterized *E.*
10 *coli* F-factor, a low copy plasmid that exists in a supercoiled circular form in host cells. The structural features of the F-factor allow stable maintenance of individual human DNA clones as well as easy manipulation of the cloned DNA. See Shizuya *et al.*, *Keio J. Med.* 50(1):26-30 (2001); Shizuya *et*
15 *al.*, *Proc. Natl. Acad. Sci. USA* 89(18):8794-7 (1992).

YACs are based on yeast linear plasmids, denoted YLp, containing homologous or heterologous DNA sequences that function as telomeres (TEL) *in vivo*, as well as containing yeast ARS (origins of replication) and CEN (centromeres)
20 segments.

HACs are human artificial chromosomes. Kuroiwa *et al.*, *Nature Biotechnol.* 18(10):1086-90 (2000); Henning *et al.*, *Proc. Natl. Acad. Sci. USA* 96(2):592-7 (1999); Harrington *et al.*, *Nature Genet.* 15(4):345-55 (1997). In one version, long
25 synthetic arrays of alpha satellite DNA are combined with telomeric DNA and genomic DNA to generate linear microchromosomes that are mitotically and cytogenetically stable in the absence of selection.

PACs are P1-derived artificial chromosomes.
30 Sternberg, *Proc. Natl. Acad. Sci. USA* 87(1):103-7 (1990); Sternberg *et al.*, *New Biol.* 2(2):151-62 (1990); Pierce *et al.*, *Proc. Natl Acad. Sci. USA* 89(6):2056-60 (1992).

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Expression vectors of the present invention - that is, those vectors that will drive expression of polypeptides from the inserted heterologous nucleic acid - will often include a variety of other genetic elements operatively linked to the protein-encoding heterologous nucleic acid insert, typically genetic elements that drive transcription, such as promoters and enhancer elements, those that facilitate RNA processing, such as transcription termination and/or polyadenylation signals, and those that facilitate translation, such as ribosomal consensus sequences.

For example, vectors for expressing proteins of the present invention in prokaryotic cells, typically *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), or the araBAD operon. Often, such prokaryotic expression vectors will further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer et al., *Proc. Natl. Acad. Sci. USA* 83:8506-8510 (1986).

As another example, vectors for expressing proteins of the present invention in yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1

promoter, the GAL1 promoter, ADH1 promoter, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

5 As another example, vectors for expressing proteins of the present invention in mammalian cells will include a promoter active in mammalian cells. Such promoters are often drawn from mammalian viruses - such as the enhancer-promoter sequences from the immediate early gene of the human
10 cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), and the enhancer-promoter from SV40. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and
15 transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

 Vector-drive protein expression can be constitutive
20 or inducible.

 Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by
25 dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PLtetO-1 promoter takes advantage of the high expression levels
30 from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in

response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline.

As another example of inducible elements, hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

Expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization.

For example, proteins of the present invention can be expressed with a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON[™] resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). As another example, the fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT[™] system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion to glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting

purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione.

Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides larger than purification and/or identification tags. Useful protein fusions include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusions for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, e.g., the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. See Barbas et al., Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001) (ISBN 0-87969-546-3); Kay et al. (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, San Diego: Academic Press, Inc., 1996; Abelson et al. (eds.), Combinatorial

Chemistry, Methods in Enzymology vol. 267, Academic Press (May 1996).

Vectors for yeast display, e.g. the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the
5 α -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, e.g., the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane
10 anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically
15 fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. These proteins are intrinsically fluorescent: the GFP-like chromophore is entirely encoded by its amino acid sequence and can fluoresce without requirement for cofactor or substrate.

Structurally, the GFP-like chromophore comprises an
20 11-stranded β -barrel (β -can) with a central α -helix, the central α -helix having a conjugated π -resonance system that includes two aromatic ring systems and the bridge between them.

The π -resonance system is created by autocatalytic cyclization among amino acids; cyclization proceeds through an
25 imidazolinone intermediate, with subsequent dehydrogenation by molecular oxygen at the C α -C β bond of a participating tyrosine.

The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such
30 as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595

(AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain
5 required for fluorescence are known in the art. Li et al., "Deletions of the *Aequorea victoria* Green Fluorescent Protein Define the Minimal Domain Required for Fluorescence," *J. Biol. Chem.* 272:28545-28549 (1997).

Alternatively, the GFP-like chromophore can be
10 selected from GFP-like chromophores modified from those found in nature. Typically, such modifications are made to improve recombinant production in heterologous expression systems (with or without change in protein sequence), to alter the excitation and/or emission spectra of the native protein, to facilitate
15 purification, to facilitate or as a consequence of cloning, or are a fortuitous consequence of research investigation.

The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the
20 art. Early results of these efforts are reviewed in Heim et al., *Curr. Biol.* 6:178-182 (1996), incorporated herein by reference in its entirety; a more recent review, with tabulation of useful mutations, is found in Palm et al., "Spectral Variants of Green Fluorescent Protein," in Green
25 Fluorescent Proteins, Conn (ed.), *Methods Enzymol.* vol. 302, pp. 378 - 394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention.

30 For example, EGFP ("enhanced GFP"), Cormack et al., *Gene* 173:33-38 (1996); U.S. Pat. Nos. 6,090,919 and 5,804,387, is a red-shifted, human codon-optimized variant of GFP that has been engineered for brighter fluorescence, higher expression in

mammalian cells, and for an excitation spectrum optimized for use in flow cytometers. EGFP can usefully contribute a GFP-like chromophore to the fusion proteins of the present invention. A variety of EGFP vectors, both plasmid and viral, are available commercially (Clontech Labs, Palo Alto, CA, USA), including vectors for bacterial expression, vectors for N-terminal protein fusion expression, vectors for expression of C-terminal protein fusions, and for bicistronic expression.

Toward the other end of the emission spectrum, EBFP ("enhanced blue fluorescent protein") and BFP2 contain four amino acid substitutions that shift the emission from green to blue, enhance the brightness of fluorescence and improve solubility of the protein, Heim *et al.*, *Curr. Biol.* 6:178-182 (1996); Cormack *et al.*, *Gene* 173:33-38 (1996). EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria; as is further discussed below, the host cell of production does not affect the utility of the resulting fusion protein. The GFP-like chromophores from EBFP and BFP2 can usefully be included in the fusion proteins of the present invention, and vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA).

Analogously, EYFP ("enhanced yellow fluorescent protein"), also available from Clontech Labs, contains four amino acid substitutions, different from EBFP, Ormö *et al.*, *Science* 273:1392-1395 (1996), that shift the emission from green to yellowish-green. Citrine, an improved yellow fluorescent protein mutant, is described in Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97:11996-12001 (2000). ECFP ("enhanced cyan fluorescent protein") (Clontech Labs, Palo Alto, CA, USA) contains six amino acid substitutions, one of which shifts the emission spectrum from green to cyan. Heim *et al.*, *Curr. Biol.* 6:178-182 (1996); Miyawaki *et al.*, *Nature*

388:882-887 (1997). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Pat. Nos. 5 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. See also 10 Conn (ed.), Green Fluorescent Protein, *Methods in Enzymol.* Vol. 302, pp 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention.

15 Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in international patent application nos. WO 97/43316, WO 97/34631, WO 96/32478, 20 WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is particularly useful.

25 Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection for integrants.

For example, the pUB6/V5-His A, B, and C vectors (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of 30 mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the

CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically
5 derived from Moloney murine leukemia virus, prove particularly useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines - such as RetroPackTM PT 67, EcoPack2TM -
10 293, AmphiPack-293, GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA) - allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus. Retroviral vectors are available with
15 a variety of selectable markers, such as resistance to neomycin, hygromycin, and puromycin, permitting ready selection of stable integrants.

The present invention further includes host cells comprising the vectors of the present invention, either present
20 episomally within the cell or integrated, in whole or in part, into the host cell chromosome.

Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired
25 fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide human STTK proteins with such post-translational
30 modifications.

As noted earlier, host cells can be prokaryotic or eukaryotic. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E.*

coli, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera*

5 *frugiperda* - e.g., Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA) - *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include COS1 and COS7 cells, chinese hamster ovary (CHO)
10 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562, Jurkat cells, and BW5147. Other mammalian cell lines are well known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA,
15 USA) and the National Institute of General medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA).

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well known in
20 the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

For example, phage lambda vectors will typically be packaged using a packaging extract (e.g., Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus
25 used to infect *E. coli*. Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells.

E. coli cells can be rendered chemically competent by treatment, e.g., with CaCl₂, or a solution of Mg²⁺, Mn²⁺, Ca²⁺,
30 Rb⁺ or K⁺, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (e.g.,

Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)).

5 Bacterial cells can be rendered electrocompetent – that is, competent to take up exogenous DNA by electroporation – by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is
10 provided online in Electroprotocols (BioRad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion.

15 Spheroplasts are prepared by the action of hydrolytic enzymes – a snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus* – to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the
20 spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca²⁺. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol. For lithium-mediated
25 transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective
30 medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl et al., *Curr. Genet.* 16(5-6):339-46 (1989). For electroporation, freshly-grown

yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker et al., *Methods Enzymol.* 194:182-7 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means.

For chemical transfection, DNA can be coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhosTM Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINETM 2000, LIPOFECTAMINETM Reagent, CELLFECTIN[®] Reagent, and LIPOFECTIN[®] Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), EffecteneTM, PolyFect[®], Superfect[®] (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf). See also, Norton et al. (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000) (ISBN 1-881299-34-1), incorporated herein by reference in its entirety.

Other transfection techniques include transfection by particle embardment. See, e.g., Cheng et al., *Proc. Natl. Acad. Sci. USA* 90(10):4455-9 (1993); Yang et al., *Proc. Natl. Acad. Sci. USA* 87(24):9568-72 (1990).

5

PROTEINS

In another aspect, the present invention provides human STTK proteins, various fragments thereof suitable for use as antigens (e.g., for epitope mapping) and for use as immunogens (e.g., for raising antibodies or as vaccines), fusions of STTK polypeptides and fragments to heterologous polypeptides, and conjugates of the proteins, fragments, and fusions of the present invention to other moieties (e.g., to carrier proteins, to fluorophores).

FIG. 3 presents the predicted amino acid sequences encoded by the human STTK cDNA clone. The amino acid sequence is further presented in SEQ ID NO: 3.

Unless otherwise indicated, amino acid sequences of the proteins of the present invention were determined as a predicted translation from a nucleic acid sequence. Accordingly, any amino acid sequence presented herein may contain errors due to errors in the nucleic acid sequence, as described in detail above. Furthermore, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes - more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409:860 - 921 (2001) - and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Small deletions and insertions can often be found that do not alter the function of the protein.

Accordingly, it is an aspect of the present invention to provide proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins at least about 65% identical in sequence to those described with particularity herein, typically at least about 70%, 75%, 80%, 85%, or 90% identical in sequence to those described with particularity herein, usefully at least about 91%, 92%, 93%, 94%, or 95% identical in sequence to those described with particularity herein, usefully at least about 96%, 97%, 98%, or 99% identical in sequence to those described with particularity herein, and, most conservatively, at least about 99.5%, 99.6%, 99.7%, 99.8% and 99.9% identical in sequence to those described with particularity herein. These sequence variants can be naturally occurring or can result from human intervention by way of random or directed mutagenesis.

For purposes herein, percent identity of two amino acid sequences is determined using the procedure of Tatiana et al., "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", *FEMS Microbiol Lett.* 174:247-250 (1999), which procedure is effectuated by the computer program BLAST 2 SEQUENCES, available online at

<http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>,
To assess percent identity of amino acid sequences, the BLASTP module of BLAST 2 SEQUENCES is used with default values of (i) BLOSUM62 matrix, Henikoff et al., *Proc. Natl. Acad. Sci USA* 89(22):10915-9 (1992); (ii) open gap 11 and extension gap 1 penalties; and (iii) gap x_dropoff 50 expect 10 word size 3 filter, and both sequences are entered in their entirety.

As is well known, amino acid substitutions occur frequently among natural allelic variants, with conservative substitutions often occasioning only *de minimis* change in protein function.

Accordingly, it is an aspect of the present invention to provide proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins having the sequence of STTK proteins, or portions thereof, with conservative amino acid substitutions. It is a further aspect to provide isolated proteins having the sequence of STTK proteins, and portions thereof, with moderately conservative amino acid substitutions. These conservatively-substituted and moderately conservatively-substituted variants can be naturally occurring or can result from human intervention.

Although there are a variety of metrics for calling conservative amino acid substitutions, based primarily on either observed changes among evolutionarily related proteins or on predicted chemical similarity, for purposes herein a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix reproduced herein below (see Gonnet et al., *Science* 256(5062):1443-5 (1992)):

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	2	-1	0	0	0	0	0	0	-1	-1	-1	0	-1	-2	0	1	1	-4	-2	0
R	-1	5	0	0	-2	2	0	-1	1	-2	-2	3	-2	-3	-1	0	0	-2	-2	-2
N	0	0	4	2	-2	1	1	0	1	-3	-3	1	-2	-3	-1	1	0	-4	-1	-2
D	0	0	2	5	-3	1	3	0	0	-4	-4	0	-3	-4	-1	0	0	-5	-3	-3
C	0	-2	-2	-3	12	-2	-3	-2	-1	-1	-2	-3	-1	-1	-3	0	0	-1	0	0
Q	0	2	1	1	-2	3	2	-1	1	-2	-2	2	-1	-3	0	0	0	-3	-2	-2
E	0	0	1	3	-3	2	4	-1	0	-3	-3	1	-2	-4	0	0	0	-4	-3	-2
G	0	-1	0	0	-2	-1	-1	7	-1	-4	-4	-1	-4	-5	-2	0	-1	-4	-4	-3
H	-1	1	1	0	-1	1	0	-1	6	-2	-2	1	-1	0	-1	0	0	-1	2	-2
I	-1	-2	-3	-4	-1	-2	-3	-4	-2	4	3	-2	2	1	-3	-2	-1	-2	-1	3
L	-1	-2	-3	-4	-2	-2	-3	-4	-2	3	4	-2	3	2	-2	-2	-1	-1	0	2
K	0	3	1	0	-3	2	1	-1	1	-2	-2	3	-1	-3	-1	0	0	-4	-2	-2
M	-1	-2	-2	-3	-1	-1	-2	-4	-1	2	3	-1	4	2	-2	-1	-1	-1	0	2
F	-2	-3	-3	-4	-1	-3	-4	-5	0	1	2	-3	2	7	-4	-3	-2	4	5	0
P	0	-1	-1	-1	-3	0	0	-2	-1	-3	-2	-1	-2	-4	8	0	0	-5	-3	-2
S	1	0	1	0	0	0	0	0	0	-2	-2	0	-1	-3	0	2	2	-3	-2	-1

T	1	0	0	0	0	0	0	-1	0	-1	-1	0	-1	-2	0	2	2	-4	-2	0
W	-4	-2	-4	-5	-1	-3	-4	-4	-1	-2	-1	-4	-1	4	-5	-3	-4	14	4	-3
Y	-2	-2	-1	-3	0	-2	-3	-4	2	-1	0	-2	0	5	-3	-2	-2	4	8	-1
V	0	-2	-2	-3	0	-2	-2	-3	-2	3	2	-2	2	0	-2	-1	0	-3	-1	3

5

For purposes herein, a "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix reproduced herein above.

As is also well known in the art, relatedness of
 10 proteins can also be characterized using a functional test, the ability of the encoding nucleic acids to base-pair to one another at defined hybridization stringencies.

It is, therefore, another aspect of the invention to provide isolated proteins not only identical in sequence to
 15 those described with particularity herein, but also to provide isolated proteins ("hybridization related proteins") that are encoded by nucleic acids that hybridize under high stringency conditions (as defined herein above) to all or to a portion of various of the isolated nucleic acids of the present invention
 20 ("reference nucleic acids"). It is a further aspect of the invention to provide isolated proteins ("hybridization related proteins") that are encoded by nucleic acids that hybridize under moderate stringency conditions (as defined herein above) to all or to a portion of various of the isolated nucleic acids
 25 of the present invention ("reference nucleic acids").

The hybridization related proteins can be alternative isoforms, homologues, paralogues, and orthologues of the STTK protein of the present invention. Particularly useful
 30 orthologues are those from other primate species, such as chimpanzee, rhesus macaque monkey, baboon, orangutan, and gorilla, from rodents, such as rats, mice, guinea pigs; from lagomorphs, such as rabbits, and from domestic livestock, such as cow, pig, sheep, horse, and goat.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody.

5 It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a
10 portion of various of the isolated STTK proteins of the present invention ("reference proteins"). Such competitive inhibition can readily be determined using immunoassays well known in the art.

 Among the proteins of the present invention that
15 differ in amino acid sequence from those described with particularity herein - including those that have deletions and insertions causing up to 10% non-identity, those having conservative or moderately conservative substitutions, hybridization related proteins, and cross-reactive proteins -
20 those that substantially retain one or more STTK activities are particularly useful. As described above, those activities include protein kinase activity.

 Residues that are tolerant of change while retaining function can be identified by altering the protein at known
25 residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham et al., *Science* 244(4908):1081-5 (1989); transposon linker scanning mutagenesis, Chen et al., *Gene* 263(1-2):39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin
30 et al., *J. Mol. Biol.* 226(3):851-65 (1992); combinatorial alanine scanning, Weiss et al., *Proc. Natl. Acad. Sci USA* 97(16):8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England

Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TNTM
In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre
Technologies Corporation, Madison, WI, USA).

As further described below, the isolated proteins of
5 the present invention can readily be used as specific
immunogens to raise antibodies that specifically recognize STTK
proteins, their isoforms, homologues, paralogues, and/or
orthologues. The antibodies, in turn, can be used, *inter alia*,
specifically to assay for the STTK proteins of the present
10 invention — e.g. by ELISA for detection of protein fluid
samples, such as serum, by immunohistochemistry or laser
scanning cytometry, for detection of protein in tissue samples,
or by flow cytometry, for detection of intracellular protein in
cell suspensions — for specific antibody-mediated isolation
15 and/or purification of STTK proteins, as for example by
immunoprecipitation, and for use as specific agonists or
antagonists of STTK action.

The isolated proteins of the present invention are
also immediately available for use as specific standards in
20 assays used to determine the concentration and/or amount
specifically of the STTK proteins of the present invention. As
is well known, ELISA kits for detection and quantitation of
protein analytes typically include isolated and purified
protein of known concentration for use as a measurement
25 standard (e.g., the human interferon- γ OptEIA kit, catalog no.
555142, Pharmingen, San Diego, CA, USA includes human
recombinant gamma interferon, baculovirus produced).

The isolated proteins of the present invention are
also immediately available for use as specific biomolecule
30 capture probes for surface-enhanced laser desorption ionization
(SELDI) detection of protein-protein interactions, WO 98/59362;
WO 98/59360; WO 98/59361; and Merchant *et al.*, *Electrophoresis*
21(6):1164-77 (2000), the disclosures of which are incorporated

herein by reference in their entireties. Analogously, the isolated proteins of the present invention are also immediately available for use as specific biomolecule capture probes on BIACORE surface plasmon resonance probes. . See Weinberger et
5 al., *Pharmacogenomics* 1(4): 395-416 (2000); Malmqvist, *Biochem. Soc. Trans.* 27(2): 335-40 (1999).

The isolated proteins of the present invention are also useful as a therapeutic supplement in patients having a specific deficiency in STTK production.

10 In another aspect, the invention also provides fragments of various of the proteins of the present invention.

The protein fragments are useful, *inter alia*, as antigenic and immunogenic fragments of STTK.

By "fragments" of a protein is here intended isolated
15 proteins (equally, polypeptides, peptides, oligopeptides), however obtained, that have an amino acid sequence identical to a portion of the reference amino acid sequence, which portion is at least 6 amino acids and less than the entirety of the reference nucleic acid. As so defined, "fragments" need not be
20 obtained by physical fragmentation of the reference protein, although such provenance is not thereby precluded.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. See, e.g., Geysen et al., "Use of peptide synthesis
25 to probe viral antigens for epitopes to a resolution of a single amino acid," *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984) and U.S. Pat. Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be
30 immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, have utility as immunogens for raising antibodies that recognize the proteins of the present invention. See, e.g., Lerner, "Tapping the immunological repertoire to produce antibodies of predetermined specificity," *Nature* 299:592-596 (1982); Shinnick et al., "Synthetic peptide immunogens as vaccines," *Annu. Rev. Microbiol.* 37:425-46 (1983); Sutcliffe et al., "Antibodies that react with predetermined sites on proteins," *Science* 219:660-6 (1983), the disclosures of which are incorporated herein by reference in their entirety. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic - that is, prove capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Pat. Nos. 5,539,084 and 5,783,674, incorporated herein by reference in their entirety.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein or the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger

fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

The present invention further provides fusions of each of the proteins and protein fragments of the present invention to heterologous polypeptides.

By fusion is here intended that the protein or protein fragment of the present invention is linearly contiguous to the heterologous polypeptide in a peptide-bonded polymer of amino acids or amino acid analogues; by "heterologous polypeptide" is here intended a polypeptide that does not naturally occur in contiguity with the protein or protein fragment of the present invention. As so defined, the fusion can consist entirely of a plurality of fragments of the STTK protein in altered arrangement; in such case, any of the STTK fragments can be considered heterologous to the other STTK fragments in the fusion protein. More typically, however, the heterologous polypeptide is not drawn from the STTK protein itself.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc

region, and even entire proteins (such as GFP chromophore-containing proteins), have particular utility.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated herein by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of STTK presence.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. See Bartel et al. (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997) (ISBN: 0195109384); Zhu et al., Yeast Hybrid Technologies, Eaton Publishing, (2000) (ISBN 1-881299-15-5); Fields et al., *Trends Genet.* 10(8):286-92 (1994); Mendelsohn et al., *Curr. Opin. Biotechnol.* 5(5):482-6 (1994); Luban et al., *Curr. Opin. Biotechnol.* 6(1):59-64 (1995); Allen et al., *Trends Biochem. Sci.* 20(12):511-6 (1995); Drees, *Curr. Opin. Chem.*

Biol. 3(1):64-70 (1999); Topcu et al., *Pharm. Res.* 17(9):1049-55 (2000); Fashena et al., *Gene* 250(1-2):1-14 (2000), the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to
5 either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful protein fusions include those that permit display of the encoded protein on the surface of a phage
10 or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The proteins and protein fragments of the present
15 invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

20 The isolated proteins, protein fragments, and protein fusions of the present invention can be composed of natural amino acids linked by native peptide bonds, or can contain any or all of nonnatural amino acid analogues, nonnative bonds, and post-synthetic (post translational) modifications, either
25 throughout the length of the protein or localized to one or more portions thereof.

As is well known in the art, when the isolated protein is used, e.g., for epitope mapping, the range of such nonnatural analogues, nonnative inter-residue bonds, or post-
30 synthesis modifications will be limited to those that permit binding of the peptide to antibodies. When used as an immunogen for the preparation of antibodies in a non-human host, such as a mouse, the range of such nonnatural analogues,

nonnative inter-residue bonds, or post-synthesis modifications will be limited to those that do not interfere with the immunogenicity of the protein. When the isolated protein is used as a therapeutic agent, such as a vaccine or for replacement therapy, the range of such changes will be limited to those that do not confer toxicity upon the isolated protein.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common.

Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan et al. (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000) (ISBN: 0199637245); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (August 1992) (ISBN: 0198556683); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (December 1993) (ISBN: 0387564314), the disclosures of which are incorporated herein by reference in their entireties.

For example, D-enantiomers of natural amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-enantiomers can also be used to confer specific three dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (Kole et al., *Biochem. Biophys. Res. Com.* 209:817-821 (1995)), and various halogenated phenylalanine derivatives.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide a labeled polypeptide.

5 Biotin, for example (indirectly detectable through interaction with avidin, streptavidin, neutravidin, captavidin, or anti-biotin antibody), can be added using biotinoyl-- (9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). (Biotin can also be added enzymatically by incorporation into a fusion protein of a *E.*
10 *coli* BirA substrate peptide.)

The FMOC and tBOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene
15 derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS--FMOC-L-glutamic acid or the corresponding tBOC derivative (both from Molecular Probes,
20 Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)--TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be
25 incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated
30 peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, e.g., Fmoc-2-

aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-
 aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-
 aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-
 amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-
 5 exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-
 cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-
 cyclohexanecarboxylic acid, Fmoc-1-amino-1-
 cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-
 cyclopentanecarboxylic acid, Fmoc-1-amino-1-
 10 cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-
 (ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric
 acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-
 aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic
 acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-
 15 carboxylic acid, Fmoc-N-(4-aminobenzoyl)-b-alanine, Fmoc-2-
 amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid,
 Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-
 hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-
 4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic
 20 acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-
 methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-
 2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic
 acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-
 methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-
 25 amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-
 D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-
 amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-?-amino-
 2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-
 4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine,
 30 Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-
 tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-
 carboxylic acid, all available from The Peptide Laboratory
 (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid and.

5 Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG
10 codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9):4780-5 (1999); Wang *et al.*, *Science* 292(5516):498-500 (2001).

The isolated proteins, protein fragments and fusion
15 proteins of the present invention can also include nonnative inter-residue bonds, including bonds that lead to circular and branched forms.

The isolated proteins and protein fragments of the present invention can also include post-translational and post-
20 synthetic modifications, either throughout the length of the protein or localized to one or more portions thereof.

For example, when produced by recombinant expression in eukaryotic cells, the isolated proteins, fragments, and fusion proteins of the present invention will typically include
25 N-linked and/or O-linked glycosylation, the pattern of which will reflect both the availability of glycosylation sites on the protein sequence and the identity of the host cell. Further modification of glycosylation pattern can be performed enzymatically.

30 As another example, recombinant polypeptides of the invention may also include an initial modified methionine residue, in some cases resulting from host-mediated processes.

When the proteins, protein fragments, and protein fusions of the present invention are produced by chemical synthesis, post-synthetic modification can be performed before deprotection and cleavage from the resin or after deprotection and cleavage. Modification before deprotection and cleavage of the synthesized protein often allows greater control, e.g. by allowing targeting of the modifying moiety to the N-terminus of a resin-bound synthetic peptide.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores.

A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), e.g., offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue,

Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

5 The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents.

 Common homobifunctional reagents include, e.g., APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3,
10 BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers
15 include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP,
20 Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

 The proteins, protein fragments, and protein fusions of the present invention can be conjugated, using such cross-
25 linking reagents, to fluorophores that are not amine- or thiol-reactive.

 Other labels that usefully can be conjugated to the proteins, protein fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast
30 reagents, and MRI contrast agents.

 The proteins, protein fragments, and protein fusions of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine

thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-STTK antibodies.

The proteins, protein fragments, and protein fusions of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half life of proteins administered intravenously for replacement therapy. Delgado et al., *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4):249-304 (1992); Scott et al., *Curr. Pharm. Des.* 4(6):423-38 (1998); DeSantis et al., *Curr. Opin. Biotechnol.* 10(4):324-30 (1999), incorporated herein by reference in their entirety. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

The isolated proteins of the present invention, including fusions thereof, can be produced by recombinant expression, typically using the expression vectors of the present invention as above-described or, if fewer than about 100 amino acids, by chemical synthesis (typically, solid phase synthesis), and, on occasion, by *in vitro* translation.

Production of the isolated proteins of the present invention can optionally be followed by purification. Purification of recombinantly expressed proteins is now well within the skill in the art. See, e.g., Thorner et al. (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Volume 326), Academic Press (2000), (ISBN: 0121822273); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001) (ISBN: 0195132947); Marshak et al., Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press

(1996) (ISBN: 0-87969-385-1); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

5 Briefly, however, if purification tags have been fused through use of an expression vector that appends such tag, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other
10 techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Purification of chemically-synthesized peptides can
15 readily be effected, e.g., by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form.

A purified protein of the present invention is an
20 isolated protein, as above described, that is present at a concentration of at least 95%, as measured on a weight basis (w/w) with respect to total protein in a composition. Such purities can often be obtained during chemical synthesis without further purification, as, e.g., by HPLC. Purified
25 proteins of the present invention can be present at a concentration (measured on a weight basis with respect to total protein in a composition) of 96%, 97%, 98%, and even 99%. The proteins of the present invention can even be present at levels of 99.5%, 99.6%, and even 99.7%, 99.8%, or even 99.9% following
30 purification, as by HPLC.

Although high levels of purity are particularly useful when the isolated proteins of the present invention are used as therapeutic agents - such as vaccines, or for

replacement therapy - the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

5 Thus, in another aspect, the present invention provides the isolated proteins of the present invention in substantially purified form. A "substantially purified protein" of the present invention is an isolated protein, as above described, present at a concentration of at least 70%,
10 measured on a weight basis with respect to total protein in a composition. Usefully, the substantially purified protein is present at a concentration, measured on a weight basis with respect to total protein in a composition, of at least 75%, 80%, or even at least 85%, 90%, 91%, 92%, 93%, 94%, 94.5% or
15 even at least 94.9%.

 In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

20 The proteins, fragments, and fusions of the present invention can usefully be attached to a substrate. The substrate can porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

 For example, the proteins, fragments, and fusions of
25 the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to
30 detect and quantify antibodies, e.g. in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the proteins, fragments, and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, e.g. in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in standard microtiter dish, the plastic is typically polystyrene.

The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction therebetween. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction therebetween.

Human STTK Proteins

In a first series of protein embodiments, the invention provides an isolated human STTK polypeptide having an amino acid sequence in SEQ ID NO: 3, which is full length STTK protein. When used as immunogens, the full length proteins of

the present invention can be used, *inter alia*, to elicit antibodies that bind to a variety of epitopes of the STTK protein.

5 The invention further provides fragments of the above-described polypeptides, particularly fragments having at least 6 amino acids, typically at least 8 amino acids, often at least 15 amino acids, and even the entirety of the sequence given in SEQ ID NO: 3.

10 The invention further provides fragments of at least 6 amino acids, typically at least 8 amino acids, often at least 15 amino acids, and even the entirety of the sequence given in SEQ ID NO: 31.

15 As described above, the invention further provides proteins that differ in sequence from those described with particularity in the above-referenced SEQ ID NOs., whether by way of insertion or deletion, by way of conservative or moderately conservative substitutions, as hybridization related proteins, or as cross-hybridizing proteins, with those that substantially retain a human STTK activity particularly useful.

20 The invention further provides fusions of the proteins and protein fragments herein described to heterologous polypeptides.

ANTIBODIES AND ANTIBODY-PRODUCING CELLS

25 In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to human STTK proteins and protein fragments of the present invention or to one or more of the proteins and protein fragments encoded by the isolated STTK nucleic acids of
30 the present invention. The antibodies of the present invention can be specific for all of linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or

protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, e.g., by solubilization in SDS.

In other embodiments, the invention provides
5 antibodies, including fragments and derivatives thereof, the binding of which can be competitively inhibited by one or more of the STTK proteins and protein fragments of the present invention, or by one or more of the proteins and protein fragments encoded by the isolated STTK nucleic acids of the
10 present invention.

As used herein, the term "antibody" refers to a polypeptide, at least a portion of which is encoded by at least one immunoglobulin gene, which can bind specifically to a first molecular species, and to fragments or derivatives thereof that
15 remain capable of such specific binding.

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are
20 admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

As is well known in the art, the degree to which an antibody can discriminate as among molecular species in a
25 mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non- STTK proteins by at least two-fold, more typically by at least 5-fold, typically by more than
30 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is

sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human adult liver, bone marrow, brain, colon, fetal liver, heart, kidney, lung, placenta, and skeletal muscle as well as a cell line HeLa.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, usefully at least about 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, and 1×10^{-10} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, and IgA, from any mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In such case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patent Nos. 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are

incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often
5 preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from
10 another species, such as mouse.

IgG, IgM, IgD, IgE and IgA antibodies of the present invention are also usefully obtained from other mammalian species, including rodents - typically mouse, but also rat, guinea pig, and hamster - lagomorphs, typically rabbits, and
15 also larger mammals, such as sheep, goats, cows, and horses. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein
20 or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin,
25 keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the
30 proteins and protein fragments of the present invention to other moieties.

For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core

matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam et al., Proc. Natl. Acad. Sci. USA 85:5409-5413 (1988); Posnett et al., J. Biol. Chem. 263, 1719-1725 (1988).

Protocols for immunizing non-human mammals are well-established in the art, Harlow et al. (eds.), Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998) (ISBN: 0879693142); Coligan et al. (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001) (ISBN: 0-471-52276-7); Zola, Monoclonal Antibodies : Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000) (ISBN: 0387915907), the disclosures of which are incorporated herein by reference, and often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant.

Antibodies from nonhuman mammals can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention.

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well known in the art, Coligan et al. (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001) (ISBN: 0-471-52276-7); Zola, Monoclonal Antibodies : Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000) (ISBN: 0387915907); Howard et al. (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000) (ISBN: 0849394457); Harlow

et al. (eds.), Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998) (ISBN: 0879693142); Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995) (ISBN: 0896033082); Delves (ed.), Antibody Production:

5 Essential Techniques, John Wiley & Son Ltd (1997) (ISBN: 0471970107); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997) (ISBN: 0412141914), incorporated herein by reference in their entirety, and thus need not be detailed here.

10 Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not
15 mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: e.g., genes encoding antibodies specific for the proteins and
20 protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S. Pat. No. 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

25 Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant antibody production – either whole antibodies, antibody fragments, or antibody
30 derivatives – can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established, Sidhu, *Curr. Opin. Biotechnol.* 11(6):610-6 (2000); Griffiths et al., *Curr. Opin. Biotechnol.* 9(1):102-8 (1998); Hoogenboom et al., *Immunotechnology*, 4(1):1-20 (1998); Rader et al., *Current Opinion in Biotechnology* 8:503-508 (1997); Aujame et al., *Human Antibodies* 8:155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15:62-70 (1997); de Kruif et al., 17:453-455 (1996); Barbas et al., *Trends in Biotechnol.* 14:230-234 (1996); Winter et al., *Ann. Rev. Immunol.* 433-455 (1994), and techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled, Barbas et al., Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001) (ISBN 0-87969-546-3); Kay et al. (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc. (1996); Abelson et al. (eds.), Combinatorial Chemistry, Methods in Enzymology vol. 267, Academic Press (May 1996), the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris*, Takahashi et al., *Biosci. Biotechnol. Biochem.* 64(10):2138-44 (2000); Freyre et al., *J. Biotechnol.* 76(2-3):157-63 (2000); Fischer et al., *Biotechnol. Appl. Biochem.* 30 (Pt 2):117-20 (1999); Pennell et al., *Res. Immunol.* 149(6):599-603 (1998); Eldin et al., *J. Immunol. Methods.* 201(1):67-75 (1997); and in *Saccharomyces cerevisiae*, Frenken et al., *Res. Immunol.* 149(6):589-99 (1998); Shusta et al., *Nature Biotechnol.* 16(8):773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells, Li et al., *Protein Expr. Purif.* 21(1):121-8 (2001); Ailor et al., *Biotechnol. Bioeng.* 58(2-3):196-203 (1998); Hsu et al., *Biotechnol. Prog.* 13(1):96-104 (1997); Edelman et al., *Immunology* 91(1):13-9 (1997); and Nesbit et al., *J. Immunol. Methods.* 151(1-2):201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, Giddings et al., *Nature Biotechnol.* 18(11):1151-5 (2000); Gavilondo et al., *Biotechniques* 29(1):128-38 (2000); Fischer et al., *J. Biol. Regul. Homeost. Agents* 14(2):83-92 (2000); Fischer et al., *Biotechnol. Appl. Biochem.* 30 (Pt 2):113-6 (1999); Fischer et al., *Biol. Chem.* 380(7-8):825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240:119-38 (1999); and Ma et al., *Plant Physiol.* 109(2):341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the

present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), review and compare bacterial, yeast, insect and
5 mammalian expression systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk
et al., *J. Biochem. (Tokyo)*. 125(2):328-33 (1999) and Ryabova
et al., *Nature Biotechnol.* 15(1):79-84 (1997), and in the milk
10 of transgenic animals, as further described in Pollock *et al.*,
J. Immunol. Methods 231(1-2):147-57 (1999), the disclosures of
which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and
15 protein fragments of the present invention, to one or more of
the proteins and protein fragments encoded by the isolated
nucleic acids of the present invention, or the binding of which
can be competitively inhibited by one or more of the proteins
and protein fragments of the present invention or one or more
20 of the proteins and protein fragments encoded by the isolated
nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab)'₂, and single chain Fv (scFv) fragments. Other useful
fragments are described in Hudson, *Curr. Opin. Biotechnol.*
25 9(4):395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or
more of the proteins and protein fragments of the present
invention, to one or more of the proteins and protein fragments
30 encoded by the isolated nucleic acids of the present invention,
or the binding of which can be competitively inhibited by one
or more of the proteins and protein fragments of the present

invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for in vivo administration, than are unmodified antibodies from non-human mammalian species.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human.

See, e.g., U.S. Pat. No. 5,807,715; Morrison et al., *Proc. Natl. Acad. Sci USA* 81(21):6851-5 (1984); Sharon et al., *Nature* 309(5966):364-7 (1984); Takeda et al., *Nature* 314(6010):452-4 (1985), the disclosures of which are incorporated herein by reference in their entirety. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann et al., *Nature* 332(6162):323-7 (1988); Co et al., *Nature* 351(6326):501-2 (1991); U.S. Pat. Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entirety.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or

more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one
5 or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

10 For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label can usefully be an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit
15 their immunohistochemical visualization are well known, and include alkaline phosphatase, β -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG);
20 o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN);
5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BlueGal;
25 iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB);
tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and
X-Glucoside.

30 Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides,

such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. See, e.g., Thorpe et al., *Methods Enzymol.* 133:331-53 (1986); Kricka et al., *J. Immunoassay* 17(1):67-83 (1996); and Lundqvist et al., *J. Biolumin. Chemilumin.* 10(6):353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, e.g., for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647

(monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, e.g., for western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H , and ^{125}I .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb , ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{99\text{m}}\text{Tc}$, ^{97}Ru , ^{90}Y , ^{90}Sr , ^{88}Y , ^{72}Se , ^{67}Cu , or ^{47}Sc .

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer et al., *Radiology* 207(2):529-38 (1998), or by radioisotopic labeling

As would be understood, use of the labels described above is not restricted to the application as for which they were mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present

5 invention. Commonly, the antibody in such immunotoxins is conjugated to Pseudomonas exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, Vol 166), Humana Press (2000) (ISBN:0896037754); and
10 Frankel et al. (eds.), Clinical Applications of Immunotoxins, Springer-Verlag New York, Incorporated (1998) (ISBN:3540640975), the disclosures of which are incorporated herein by reference in their entirety, for review.

The antibodies of the present invention can usefully
15 be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic
20 acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a
25 substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-
30 activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres,

typically by biotin-streptavidin interaction, which microsphere can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells.

It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Human STTK Antibodies

In a first series of antibody embodiments, the invention provides antibodies, both polyclonal and monoclonal, and fragments and derivatives thereof, that bind specifically to a polypeptide having an amino acid in SEQ ID NO: 3, which are full length human STTK proteins.

In a second series of antibody embodiments, the invention provides antibodies, both polyclonal and monoclonal, and fragments and derivatives thereof, that bind specifically

to a polypeptide having an amino acid sequence in SEQ ID NO: 31 which is a novel portion of the STTK protein.

Such antibodies are useful in *in vitro* immunoassays, such as ELISA, western blot or immunohistochemical assay of disease cells and tissues. Such antibodies are also useful in isolating and purifying STTK proteins, including related cross-reactive proteins, by immunoprecipitation, immunoaffinity chromatography, or magnetic bead-mediated purification.

In another series of antibody embodiments, the invention provides antibodies, both polyclonal and monoclonal, and fragments and derivatives thereof, the specific binding of which can be competitively inhibited by the isolated proteins and polypeptides of the present invention.

In other embodiments, the invention further provides the above-described antibodies detectably labeled, and in yet other embodiments, provides the above-described antibodies attached to a substrate.

PHARMACEUTICAL COMPOSITIONS

Human STTK is important for transducing signals in the cell; defects in STTK expression, activity, distribution, localization, and/or solubility are a cause of human disease, which disease can manifest as a disorder of adult liver, bone marrow, brain, colon, fetal liver, heart, kidney, lung, placenta, and skeletal muscle as well as a cell line HeLa function. Accordingly, pharmaceutical compositions comprising nucleic acids, proteins, and antibodies of the present invention, as well as mimetics, agonists, antagonists, or inhibitors of STTK activity, can be administered as therapeutics for treatment of STTK defects.

Thus, in another aspect, the invention provides pharmaceutical compositions comprising the nucleic acids,

nucleic acid fragments, proteins, protein fusions, protein fragments, antibodies, antibody derivatives, antibody fragments, mimetics, agonists, antagonists, and inhibitors of the present invention.

5 Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

10 Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000) (ISBN: 0683306472); Ansel et al., Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins Publishers (1999) (ISBN: 0683305727); and Kibbe (ed.), Handbook of Pharmaceutical
15 Excipients American Pharmaceutical Association, 3rd ed. (2000) (ISBN: 091733096X), the disclosures of which are incorporated herein by reference in their entirety, and thus need not be described in detail herein.

20 Briefly, however, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes,
25 including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

30 Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from
5 corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate,
10 dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as
15 sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl
20 methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the
25 aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout.

For example, dragee cores can be used in conjunction
30 with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide,

lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution.

Intramuscular preparations, e.g. a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (e.g., ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition.

The pharmaceutical compositions of the present invention can be administered topically.

A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, e.g., 5 to 10%, in a carrier such as a pharmaceutical cream base. Various formulations for topical use include drops, tinctures, lotions, creams, solutions, and ointments

containing the active ingredient and various supports and vehicles. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient - for example STTK protein, fusion protein, or fragments thereof, antibodies specific for STTK, agonists, antagonists or inhibitors of STTK - which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats,

rabbits, dogs, or pigs. The animal model can also be used to determine an initial useful concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are particularly useful.

The data obtained from cell culture assays and animal studies is used in formulating an initial dosage range for human use, and preferably provides a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a

protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (e.g., 1mg/kg to 5 mg/kg). The pharmaceutical formulation can
5 be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will
10 employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary
15 skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

THERAPEUTIC METHODS

The present invention further provides methods of treating subjects having defects in STTK - e.g., in expression,
25 activity, distribution, localization, and/or solubility of STTK - which can manifest as a disorder of adult liver, bone marrow, brain, colon, fetal liver, heart, kidney, lung, placenta, and skeletal muscle as well as a cell line HeLa function. As used herein, "treating" includes all medically-
30 acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease.

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a

pharmaceutical composition comprising STTK protein, fusion, fragment or derivative thereof is administered to a subject with a clinically-significant STTK defect.

Protein compositions are administered, for example, to complement a deficiency in native STTK. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to STTK. The immune response can be used to modulate activity of STTK or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate STTK.

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of STTK protein, fusion, or fragment thereof, or without such vector.

Nucleic acid compositions that can drive expression of STTK are administered, for example, to complement a deficiency in native STTK, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used - see, e.g., Cid-Arregui (ed.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co., 2000 (ISBN: 188129935X) - as can plasmids .

Antisense nucleic acid compositions, or vectors that drive expression of STTK antisense nucleic acids, are administered to downregulate transcription and/or translation of STTK in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have sequence that is complementary to coding or to noncoding regions of the STTK gene. For example, oligonucleotides derived from the transcription initiation site, e.g., between
5 positions -10 and +10 from the start site, are particularly useful.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to STTK transcripts, are also useful in therapy. See, e.g., Phylactou,
10 *Adv. Drug Deliv. Rev.* 44(2-3):97-108 (2000); Phylactou et al., *Hum. Mol. Genet.* 7(10):1649-53 (1998); Rossi, *Ciba Found. Symp.* 209:195-204 (1997); and Sigurdsson et al., *Trends Biotechnol.* 13(8):286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

15 Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the STTK genomic locus. Such triplexing oligonucleotides are able to inhibit transcription, Intody et al., *Nucleic Acids Res.* 28(21):4283-90 (2000);
20 McGuffie et al., *Cancer Res.* 60(14):3790-9 (2000), the disclosures of which are incorporated herein by reference, and pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a
25 pathophysiologic basis of disease.

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is
30 administered. As is well known, antibody compositions are administered, for example, to antagonize activity of STTK, or to target therapeutic agents to sites of STTK presence and/or accumulation.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of STTK is administered. Antagonists of STTK can be produced using methods generally known in the art. In particular, purified STTK can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of STTK.

In other embodiments a pharmaceutical composition comprising an agonist of STTK is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express STTK, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in STTK production or activity.

In other embodiments, pharmaceutical compositions comprising the STTK proteins, nucleic acids, antibodies, antagonists, and agonists of the present invention can be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy can be made by one of ordinary skill in the art according to conventional pharmaceutical principles. The combination of therapeutic agents or approaches can act additively or synergistically to effect the treatment or prevention of the various disorders described above, providing greater therapeutic efficacy and/or permitting use of the pharmaceutical compositions of the present invention using lower dosages, reducing the potential for adverse side effects.

TRANSGENIC ANIMALS AND CELLS

In another aspect, the invention provides transgenic cells and non-human organisms comprising human STTK isoform
5 nucleic acids, and transgenic cells and non-human organisms with targeted disruption of the endogenous orthologue of the STTK gene.

The cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric,
10 nonchimeric heterozygotes, and nonchimeric homozygotes.

DIAGNOSTIC METHODS

The nucleic acids of the present invention can be
15 used as nucleic acid probes to assess the levels of STTK mRNA in disease cells and tissues, and antibodies of the present invention can be used to assess the expression levels of STTK proteins in disease cells and tissues to diagnose cancer.

The following examples are offered for purpose of
20 illustration, not limitation.

EXAMPLE 1

Identification and Characterization of cDNAs Encoding Human STTK Proteins

25 Predicating our gene discovery efforts on use of genome-derived single exon probes and hybridization to genome-derived single exon microarrays – an approach that we have previously demonstrated will readily identify novel genes that
30 have proven refractory to mRNA-based identification efforts – we identified one exon in raw human genomic sequence that is particularly expressed in human adult liver, bone marrow, brain, fetal liver, kidney, lung, and placenta.

Briefly, bioinformatic algorithms were applied to human genomic sequence data to identify putative exons. Each of the predicted exons was amplified from genomic DNA, typically centering the putative coding sequence within a larger amplicon that included flanking noncoding sequence. These genome-derived single exon probes were arrayed on a support and expression of the bioinformatically predicted exons assessed through a series of simultaneous two-color hybridizations to the genome-derived single exon microarrays.

The approach and procedures are further described in detail in Penn *et al.*, "Mining the Human Genome using Microarrays of Open Reading Frames," *Nature Genetics* 26:315-318 (2000); commonly owned and copending U.S. patent application nos. 09/864,761, filed May 23, 2001, 09/774,203, filed January 29, 2001, and 09/632,366, filed August 3, 2000, the disclosures of which are incorporated herein by reference in their entireties.

Using a graphical display particularly designed to facilitate computerized query of the resulting exon-specific expression data, as further described in commonly owned and copending U.S. patent application no. 09/864,761, filed May 23, 2001, 09/774,203, filed January 29, 2001 and 09/632,366, filed August 3, 2000, the disclosures of which are incorporated herein by reference in their entireties, 3 exons were identified that are expressed in all the human tissues tested; subsequent analysis revealed that the 3 exons belong to the same gene.

Table 1 summarizes the microarray expression data obtained using genome-derived single exon probes corresponding to exons three, eight, and thirteen. Each probe was completely sequenced on both strands prior to its use on a genome-derived single exon microarray; sequencing confirmed the exact chemical structure of each probe. An added benefit of sequencing is

that it placed us in possession of a set of single base-incremented fragments of the sequenced nucleic acid, starting from the sequencing primer's 3' OH. (Since the single exon probes were first obtained by PCR amplification from genomic DNA, we were of course additionally in possession of an even larger set of single base incremented fragments of each of the single exon probes, each fragment corresponding to an extension product from one of the two amplification primers.)

Signals are normalized values measured and calculated as further described in commonly owned and copending U.S. patent application nos. 09/864,761, filed May 23, 2001, 09/774,203, filed January 29, 2001, 09/632,366, filed August 3, 2000, and U.S. provisional patent application no. 60/207,456, filed May 26, 2000, the disclosures of which are incorporated herein by reference in their entireties.

Table 1 Expression Analysis Genome-Derived Single Exon Microarray			
TISSUE	Amplicon 29277, Exon 3	Amplicon 29278, Exon 8	Amplicon 29279, Exon 13
adult liver	0.49		
bone marrow		0.43	0.77
brain	0.69	0.43	0.88
fetal liver	0.61	0.36	0.80
kidney		0.53	
lung	0.59		0.72
placenta			1.01

As shown in Table 1, significant expression of exon
exons three, eight, and thirteen was seen in adult liver, bone
marrow, brain, fetal liver, kidney, lung, and placenta. Adult
liver, bone marrow, lung and placenta-specific expression was
5 further confirmed by RT-PCR analysis (see below).

Marathon-Ready™ brain cDNA (Clontech Laboratories,
Palo Alto, CA, USA) was used as a substrate for standard RACE
(rapid amplification of cDNA ends) to obtain a cDNA clone that
spans 2.9 kilobases and appears to contain the entire coding
10 region of the gene to which the exon contributes; for reasons
described below, we termed this cDNA human STTK. Specifically,
oligonucleotides OL699 (5'-ATGGAGGGCGACGGGGTGCCATGGGGCAGCG-3';
SEQ ID NO: 32) and OL704 (5'-GCCTGGTTAATAAAGATGTCAAGGCAGTGGCAG-
3'; SEQ ID NO: 35) were used to PCR out a 1.13 kb fragment of
15 the open reading frame (ORF) using protocols according to
manufacture's instructions (Clontech Laboratories). Based on
this sequence, OL700 (5'-CAGACTCGCCTGGTGGATGCAGCCAAGGC-3'; SEQ
ID NO: 33) and OL701 (5'-CTGCCACTGCCTTGACATCTTTATTAACCAGGC-3';
SEQ ID NO: 34) were used, together with the RACE adaptor
20 primers, to RACE out the 3' end sequence. Finally,
Oligonucleotides OL701 (SEQ ID NO: 34) and OL776 (5'-
CCCATGGCCAAAAGGTGAGGGGGAAGGAAAT-3'; SEQ ID NO: 36) were used to
PCR out another 1.76 kb fragment of the ORF. The products were
sequenced and the full length ORF of human STTK gene was
25 assembled from these sequences.

The human STTK cDNA was sequenced on both strands
using a MegaBACE™ 1000 sequencer (Amersham Biosciences,
Sunnyvale, CA, USA). Sequencing both strands provided us with
the exact chemical structure of the cDNA, which is shown in
30 FIG. 3 and further presented in the SEQUENCE LISTING as SEQ ID
NO: 1, and placed us in actual physical possession of the
entire set of single-base incremented fragments of the
sequenced clone, starting at the 5' and 3' termini.

As shown in FIG. 3, the human STTK cDNA spans 2853 nucleotides and contains an open reading frame from nucleotide 1 through and including nt 2790 (inclusive of termination codon), predicting a protein of 929 amino acids with a

5 (posttranslationally unmodified) molecular weight of 105.3 kD.

The clone appears full length, with the reading frame starting with a methionine and terminating with a stop codon.

BLAST query of genomic sequence identified 1 BAC, spanning over 64 kb, that constitute the minimum set of clones
10 encompassing the cDNA sequence. Based upon the known origin of the BAC (GenBank accession numbers AC018711.4, the human STTK gene can be mapped to human chromosome 1q32.1.

Comparison of the cDNA and genomic sequences identified 13 exons. Exon organization is listed in Table 2.

Table 2 Human STTK Exon Structure			
Exon no.	cDNA range	genomic range	BAC accession
1	1-265	83364-83628	AC018711.4
2	266-654	107093-107481	
3	655-1324	125067-125736	
4	1325-1557	130944-131176	
5	1558-1641	131893-131976	
6	1642-1818	132687-132863	
7	1819-1948	133512-133641	
8	1949-2105	134629-134785	
9	2106-2238	135220-135352	
10	2239-2352	137513-137626	
11	2353-2467	144105-144219	
12	2468-2602	146560-146694	

13	2603-2853	147154-147404	
----	-----------	---------------	--

FIG. 2 schematizes the exon organization of the human STTK clone.

At the top is shown the one bacterial artificial chromosome (BAC), with GenBank accession number (AC018711.4), that spans the STTK locus. The genome-derived single-exon probes first used to demonstrate expression from this locus are shown below the BACs and labeled "516", "578", and "550", respectively. The 516 bp probe includes sequence drawn from exon 3, with additional sequence from intron 3. The 578 bp probe includes sequence drawn from exon 8, with additional sequence from intron 7 and intron 8. The 550 bp probe includes sequence drawn from exon 13, with additional sequence from intron 12 as well as sequence 3' of exon 13.

As shown in FIG. 2, STTK encodes a protein of 929 amino acids and is comprised of exons 1 - 13. STTK has a predicted molecular weights, prior to any post-translational modification, of 105.3 kD.

As further discussed in the examples herein, expression of STTK was assessed using hybridization to genome-derived single exon microarrays and RT-PCR. Microarray analysis of exons three, eight, and thirteen showed expression in adult liver, bone marrow, brain, fetal liver, kidney, lung, and placenta. RT-PCR confirmed microarray data, and further provided expression data for colon, heart, skeletal muscle as well as a cell line HeLa.

The sequence of the human STTK cDNA was used as a BLAST query into the GenBank nr and dbEst databases. The nr database includes all non-redundant GenBank coding sequence translations, sequences derived from the 3-dimensional structures in the Brookhaven Protein Data Bank (PDB), sequences

from SwissProt, sequences from the protein information resource (PIR), and sequences from protein research foundation (PRF). The dbEst (database of expressed sequence tags) includes ESTs, short, single pass read cDNA (mRNA) sequences, and cDNA
5 sequences from differential display experiments and RACE experiments.

BLAST search identified multiple human and mouse ESTs, one EST from zebrafish (AI331501.1), two ESTs from *Xenopus laevis* (AW199365.1 and AW200427.1) as having sequence
10 closely related to STTK.

Globally, the human STTK protein resembles a putative mouse gene (GenBank accession: AK020880.1, with 81 % amino acid identity and 86 % amino acid similarity over amino acids 1-394 of the human STTK protein), as well as a putative *Arabidopsis*
15 transcript (GenBank accession: AY035004.1, with 34 % amino acid identity and 52 % amino acid similarity over amino acids 737-909 of the human STTK protein).

Motif searches using Pfam (<http://pfam.wustl.edu>), SMART (<http://smart.embl-heidelberg.de>), and PROSITE pattern
20 and profile databases (<http://www.expasy.ch/prosite>), identified one known domain shared with protein kinases.

FIG. 1 shows the domain structure of human STTK.

As schematized in FIG. 1, the newly isolated gene product shares a key protein domain, the kinase domain, with
25 other protein kinases. The shared structural feature strongly implies that human STTK plays a role similar to that of other protein kinases in signal transduction in the cell.

Like other protein kinases, STTK has a protein kinase domain. In STTK, the kinase domain occurs at residues 653 - 897
30 (<http://pfam.wustl.edu/>). Protein kinase domain is the catalytic domain of a protein kinase that catalyzes the transfer of the γ -phosphate of ATP to protein substrates. Protein kinases can be divided into two major groups, the

serine/threonine kinases and the tyrosine kinases, depending on their substrate specificities. Since the kinase domain of STTK is similarly homologous to the serine/threonine kinase domain and the tyrosine kinase domain, STTK is predicted to be able to phosphorylate serine/threonine as well as tyrosine residues in its substrates.

Other signatures of the STTK protein were identified by searching the PROSITE database (<http://www.expasy.ch/tools/scnpsit1.html>). These include two N-glycosylation sites (60-63, 594-597), fifteen Protein kinase C phosphorylation sites, nine Casein kinase II phosphorylation sites, two Tyrosine kinase phosphorylation site (305-312 and 395-402), and fifteen N-myristoylation sites.

Possession of the genomic sequence permitted search for promoter and other control sequences for the human STTK gene.

A putative transcriptional control region, inclusive of promoter and downstream elements, was defined as 1 kb around the transcription start site, itself defined as the first nucleotide of the human STTK cDNA clone. The region, drawn from sequence of BAC AC018711.4, has the sequence given in SEQ ID NO: 30, which lists 1000 nucleotides before the transcription start site.

Transcription factor binding sites were identified using a web based program (<http://motif.genome.ad.jp/>), including binding sites for CRE-BP1/c-Jun (79-86), E2F (627-634), SRY (477-483, 633-639, and 841-847), Lmo2 (816-824), v-Myb (847-855), Sp1 (932-941), and GAZA-2 (794-803 and 815-824, with numbering according to SEQ ID NO: 30), amongst others.

We have thus identified a newly described human gene, STTK, which shares the protein kinase domain with other protein kinases. The shared structural feature strongly implies that the STTK protein plays a role similar to other protein kinases,

transducing signals in the cell, making the STTK proteins and nucleic acids clinically useful diagnostic markers and potential therapeutic agents for cancer.

5

EXAMPLE 2
Preparation and Labeling of
Useful Fragments of Human STTK

10 Useful fragments of STTK are produced by PCR, using standard techniques, or solid phase chemical synthesis using an automated nucleic acid synthesizer. Each fragment is sequenced, confirming the exact chemical structure thereof.

15 The exact chemical structure of preferred fragments is provided in the attached SEQUENCE LISTING, the disclosure of which is incorporated herein by reference in its entirety. The following summary identifies the fragments whose structures are more fully described in the SEQUENCE LISTING:

20 SEQ ID NO: 1 (nt, full length STTK cDNA)
SEQ ID NO: 2 (nt, cDNA ORF)
SEQ ID NO: 3 (aa, full length STTK protein)
SEQ ID NO: 4 - 16 (nt, exons 1 - 13 (from genomic
sequence))
SEQ ID NO: 17 - 29 (nt, 500 bp genomic amplicon centered
25 about exons 1 - 13)
SEQ ID NO: 30 (nt, 1000 bp putative promoter)
SEQ ID NO: 31 (aa, (aa 333 - 539) portion of STTK)

SEQ ID NO: 32 (nt, primer OL699 for PCR cloning of STTK)
SEQ ID NO: 33 (nt, primer OL700 for PCR cloning of STTK)
SEQ ID NO: 34 (nt, primer OL701 for PCR cloning of STTK)
SEQ ID NO: 35 (nt, primer OL704 for PCR cloning of STTK)
5 SEQ ID NO: 36 (nt, primer OL776 for PCR cloning and
expression analysis of STTK)
SEQ ID NO: 37 (nt, primer OL722 for expression analysis
of STTK by RT-PCR)

10 Upon confirmation of the exact structure, each of the
above-described nucleic acids of confirmed structure is
recognized to be immediately useful as a STTK-specific probe.

For use as labeled nucleic acid probes, the above-
described STTK nucleic acids are separately labeled by random
15 priming. As is well known in the art of molecular biology,
random priming places the investigator in possession of a near-
complete set of labeled fragments of the template of varying
length and varying starting nucleotide.

The labeled probes are used to identify the STTK gene
20 on a Southern blot, and are used to measure expression of STTK
mRNA on a northern blot and by RT-PCR, using standard
techniques.

EXAMPLE 3

Expression Analysis of STTK by RT-PCR

25 To explore the potential function of the STTK gene,
the expression of STTK in human tissues was examined by PCR
30 using marathon-ready cDNAs (Clontech). Oligonucleotides OL722
(5'-ATGGAGGGCGACGGGGTGCCATGGGGCAGCGAGC-3'; SEQ ID NO: 37) and
OL776 (SEQ ID NO: 36) were used to amplify human STTK ORFs from
human cDNAs of bone marrow, brain, colon, heart, kidney, liver,
lung, placenta, skeletal muscle and HeLa cells. The PCR
35 conditions were according to a touchdown PCR procedure. The

tubes containing the oligonucleotides, cDNA and Taq polymerase were first incubated at 94°C for 15 seconds followed by 72°C for 3 minutes, cycle 5 times. The tubes were then incubated at 94°C for 15 seconds followed by 70°C for 3 minutes, cycle 5 times. Finally the tubes were incubated at 94°C for 15 seconds followed by 68°C for 3 minutes, cycle 25 times. The result of the expression profile is shown in FIG. 4. The abundance of PCR product indicates that human STTK gene is highly expressed in bone marrow, brain, colon, heart, kidney, liver, placenta, skeletal muscle. It has lower expression in HeLa cells and very low expression in lung (Figure 4).

EXAMPLE 4 Production of Human STTK Protein

The full length STTK cDNA clone is cloned into the mammalian expression vector pCDNA3.1/HISA (Invitrogen, Carlsbad, CA, USA), transfected into COS7 cells, transfectants selected with G418, and protein expression in transfectants confirmed by detection of the anti-XpressTM epitope according to manufacturer's instructions. Protein is purified using immobilized metal affinity chromatography and vector-encoded protein sequence is then removed with enterokinase, per manufacturer's instructions, followed by gel filtration and/or HPLC.

Following epitope tag removal, STTK protein is present at a concentration of at least 70%, measured on a weight basis with respect to total protein (i.e., w/w), and is free of acrylamide monomers, bis acrylamide monomers, polyacrylamide and ampholytes. Further HPLC purification provides STTK protein at a concentration of at least 95%, measured on a weight basis with respect to total protein (i.e., w/w).

EXAMPLE 5
Production of Anti-STTK Antibody

Purified proteins prepared as in Example 4 are
5 conjugated to carrier proteins and used to prepare murine
monoclonal antibodies by standard techniques. Initial
screening with the unconjugated purified proteins, followed by
competitive inhibition screening using peptide fragments of the
STTK, identifies monoclonal antibodies with specificity for
10 STTK.

EXAMPLE 6
Use of STTK Probes and Antibodies
for the Diagnosis of Cancer

15 After informed consent is obtained, samples are drawn
from disease cells or tissues and tested for STTK mRNA levels
by standard techniques and tested additionally for STTK protein
levels using anti-STTK antibodies in a standard ELISA.

EXAMPLE 7
Use of STTK Nucleic Acids, Proteins,
and Antibodies in Therapy

20 Once over-expression of STTK is detected in patients,
STTK antisense RNA or STTK-specific antibody is introduced into
disease cells to reduce the amount of the protein.

Once mutations of STTK have been detected in
patients, normal STTK is reintroduced into the patient's
disease cells by introduction of expression vectors that drive
25 STTK expression or by introducing STTK proteins into cells.
Antibodies for the mutated forms of STTK are used to block the
function of the abnormal forms of the protein.

35

EXAMPLE 8

STTK Disease Associations

Diseases that map to the human STTK chromosomal region (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>) are shown in Table 3. More recently, allelic gains of the 1q32 region have been observed in several types of cancer, including primary hepatocellular carcinoma (Wang et al., *Genes Chromosomes Cancer* 31:221-227 (2001)), breast cancer (Forozan et al., *Cancer Res.* 60:4519-4525 (2000)), and esophageal squamous cell carcinoma (Pimkhaokham et al., *Jpn. J. Cancer Res.* 91:1126-1133 (2000)). Mutations of STTK might lead to the disease(s) listed below. Alternatively, mutations of STTK might lead to some other human disorder(s) as well.

Table 3:
Diseases mapped to human chromosome 1q32.1 (STTK region).

mim_num	disease	chromosomal location
104300	ALZHEIMER DISEASE	1q31-q42
602390	HEMOCHROMATOSIS, TYPE 2	1q

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.